

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C07K 14/705, A61K 38/17, C07K 16/28, G01N 33/53</b>		<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 99/14241</b> <b>(43) International Publication Date:</b> 25 March 1999 (25.03.99)																				
<b>(21) International Application Number:</b> PCT/US98/19437 <b>(22) International Filing Date:</b> 17 September 1998 (17.09.98)			<b>(74) Agents:</b> SCHWARTZ, Timothy, R. et al.; Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080-4990 (US).  <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).																				
<b>(30) Priority Data:</b> <table border="0"><tr><td>60/059,119</td><td>17 September 1997 (17.09.97)</td><td>US</td></tr><tr><td>60/059,263</td><td>18 September 1997 (18.09.97)</td><td>US</td></tr><tr><td>60/063,550</td><td>28 October 1997 (28.10.97)</td><td>US</td></tr><tr><td>60/066,770</td><td>12 November 1997 (12.11.97)</td><td>US</td></tr><tr><td>60/066,364</td><td>21 November 1997 (21.11.97)</td><td>US</td></tr><tr><td>60/065,186</td><td>24 November 1997 (24.11.97)</td><td>US</td></tr><tr><td>60/088,026</td><td>4 June 1998 (04.06.98)</td><td>US</td></tr></table> <b>(71) Applicant (for all designated States except US):</b> GENENTECH, INC. [US/US]; 1 DNA Way, South San Francisco, CA 94080-4990 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> FONG, Sherman [US/US]; 19 Basinside Way, Alameda, CA 94502 (US). GODDARD, Audrey [CA/US]; 110 Congo Street, San Francisco, CA 94131 (US). GURNEY, Austin, L. [US/US]; 1 Debbie Lane, Belmont, CA 94002 (US). TUMAS, Daniel [US/US]; 3 Rae Avenue, Orinda, CA 94563 (US). WOOD, William, I. [US/US]; 1400 Tarrytown Street, San Mateo, CA 94402 (US).				60/059,119	17 September 1997 (17.09.97)	US	60/059,263	18 September 1997 (18.09.97)	US	60/063,550	28 October 1997 (28.10.97)	US	60/066,770	12 November 1997 (12.11.97)	US	60/066,364	21 November 1997 (21.11.97)	US	60/065,186	24 November 1997 (24.11.97)	US	60/088,026	4 June 1998 (04.06.98)
60/059,119	17 September 1997 (17.09.97)	US																					
60/059,263	18 September 1997 (18.09.97)	US																					
60/063,550	28 October 1997 (28.10.97)	US																					
60/066,770	12 November 1997 (12.11.97)	US																					
60/066,364	21 November 1997 (21.11.97)	US																					
60/065,186	24 November 1997 (24.11.97)	US																					
60/088,026	4 June 1998 (04.06.98)	US																					
<b>(54) Title:</b> COMPOSITIONS AND METHODS FOR THE TREATMENT OF IMMUNE RELATED DISEASES																							
<b>(57) Abstract</b> <p>The present invention relates to a composition containing novel proteins and methods for the diagnosis and treatment of immune related diseases.</p>																							

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## COMPOSITIONS AND METHODS FOR THE TREATMENT OF IMMUNE RELATED DISEASES

Field of the Invention

The present invention relates to compositions and methods for the diagnosis and treatment of immune related diseases.

5

Background of the Invention

Immune related and inflammatory diseases are the manifestation or consequence of fairly complex, often multiple interconnected biological pathways which in normal physiology are critical to respond to insult or injury, initiate repair from insult or injury, and mount innate and acquired defense against foreign organisms. Disease or pathology occurs when these normal physiological pathways cause additional insult or injury either as directly related to the intensity of the response, as a consequence of abnormal regulation or excessive stimulation, as a reaction to self, or as a combination of these.

Though the genesis of these diseases often involves multistep pathways and often multiple different biological systems/pathways, intervention at critical points in one or more of these pathways can have an ameliorative or therapeutic effect. Therapeutic intervention can occur by either antagonism of a detrimental process/pathway or stimulation of a beneficial process/pathway.

Many immune related diseases are known and have been extensively studied. Such diseases include immune-mediated inflammatory diseases, non-immune-mediated inflammatory diseases, infectious diseases, immunodeficiency diseases, neoplasia, etc.

T lymphocytes (T cells) are an important component of a mammalian immune response. T cells recognise antigens which are associated with a self-molecule encoded by genes within the major histocompatibility complex (MHC). The antigen may be displayed together with MHC molecules on the surface of antigen presenting cells, virus infected cells, cancer cells, grafts, etc. The T cell system eliminates these altered cells which pose a health threat to the host mammal. T cells include helper T cells and cytotoxic T cells. Helper T cells proliferate extensively following recognition of an antigen-MHC complex on an antigen presenting cell. Helper T cells also secrete a variety of cytokines, i.e. lymphokines, which play a central role in the activation of B cells, cytotoxic T cells and a variety of other cells which participate in the immune response.

A central event in both humoral and cell mediated immune responses is the activation and clonal expansion of helper T cells. Helper T cell activation is initiated by the interaction of the T cell receptor (TCR) - CD3 complex with an antigen-MHC on the surface of an antigen presenting cell. This interaction mediates a cascade of biochemical events that induce the resting helper T cell to enter a cell cycle (the G<sub>0</sub> to G<sub>1</sub> transition) and results in the expression of a high affinity receptor for IL-2 and sometimes IL-4. The activated T cell progresses through the cycle proliferating and differentiating into memory cells or effector cells.

In addition to the signals mediated through the TCR, activation of T cells involves additional costimulation induced by cytokines released by the antigen presenting cell or through interactions with membrane bound molecules on the antigen presenting cell and the T cell. The cytokines IL-1 and IL-6 have been shown to provide a costimulatory signal. Also, the interaction between the B7 molecule expressed on the surface of an antigen presenting cell and CD28 and CTLA-4 molecules expressed on the T cell surface effect T cell activation. Activated T cells express an increased number of cellular adhesion molecules, such as ICAM-1, integrins, VLA-4, LFA-1, CD56, etc.

T-cell proliferation in a mixed lymphocyte culture or mixed lymphocyte reaction (MLR) is an established indication of the ability of a compound to stimulate the immune system. In many immune

responses, inflammatory cells infiltrate the site of injury or infection. The migrating cells may be neutrophilic, eosinophilic, monocytic or lymphocytic. Histologic examination of the affected tissues provides evidence of an immune stimulating or inhibiting response. Current Protocols in Immunology, ed. John E. Coligan, 1994, John Wiley & Sons, Inc. .

- 5 Immune related diseases can be treated by suppressing the immune response. Using neutralizing antibodies that inhibit molecules having immune stimulatory activity would be beneficial in the treatment of immune-mediated and inflammatory diseases. Molecules which inhibit the immune response can be utilized (proteins directly or via the use of antibody agonists) to inhibit the immune response and thus ameliorate immune related disease.

10

#### Summary of the Invention

- The present invention concerns compositions and methods for the diagnosis and treatment of immune related disease in mammals, including humans. The present invention is based on the identification of proteins (including agonist and antagonist antibodies) which either stimulate or inhibit the immune response in mammals. Immune related diseases can be treated by suppressing or enhancing the immune response. Molecules that enhance the immune response stimulate or potentiate the immune response to an antigen. Molecules which stimulate the immune response can be used therapeutically where enhancement of the immune response would be beneficial. Such stimulatory molecules can also be inhibited where suppression of the immune response would be of value. Neutralizing antibodies are examples of molecules that inhibit molecules having immune stimulatory activity and which would be beneficial in the treatment of immune related and inflammatory diseases. Molecules which inhibit the immune response can also be utilized (proteins directly or via the use of antibody agonists) to inhibit the immune response and thus ameliorate immune related disease.

- Accordingly, the proteins of the invention encoded by the genes of the invention are useful for the diagnosis and/or treatment (including prevention) of immune related diseases. Antibodies which bind to stimulatory proteins are useful to suppress the immune system and the immune response. Antibodies which bind to inhibitory proteins are useful to stimulate the immune system and the immune response. The proteins and antibodies of the invention are also useful to prepare medicines and medicaments for the treatment of immune related and inflammatory diseases.

- In one embodiment, the present invention concerns an isolated antibody which binds a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. In one aspect, the antibody mimics the activity of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide (an agonist antibody) or conversely the antibody inhibits or neutralizes the activity of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide (an antagonist antibody). In another aspect, the antibody is a monoclonal antibody, which preferably has nonhuman complementarity determining region (CDR) residues and human framework region (FR) residues. The antibody may be labeled and may be immobilized on a solid support. In a further aspect, the antibody is an antibody fragment, a single-chain antibody, or an anti-idiotypic antibody.

- In another embodiment, the invention concerns a composition containing a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide or an agonist or antagonist antibody which binds the polypeptide in admixture with a carrier or excipient. In one aspect, the composition contains a therapeutically effective amount of the peptide or antibody. In another aspect, when the composition contains an immune stimulating molecule, the composition is useful for: (a) increasing infiltration of



inflammatory cells into a tissue of a mammal in need thereof, (b) stimulating or enhancing an immune response in a mammal in need thereof, or (c) increasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen. In a further aspect, when the composition contains an immune inhibiting molecule, the composition is useful for: (a) decreasing infiltration of inflammatory cells into a tissue of a mammal in need thereof, (b) inhibiting or reducing an immune response in a mammal in need thereof, or (c) decreasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen. In another aspect, the composition contains a further active ingredient, which may, for example, be a further antibody or a cytotoxic or chemotherapeutic agent. Preferably, the composition is sterile.

In another embodiment, the invention concerns the use of the polypeptides and antibodies of the invention to prepare a composition or medicament which has the uses described above.

In a further embodiment, the invention concerns nucleic acid encoding an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody, and vectors and recombinant host cells comprising such nucleic acid. In a still further embodiment, the invention concerns a method for producing such an antibody by culturing a host cell transformed with nucleic acid encoding the antibody under conditions such that the antibody is expressed, and recovering the antibody from the cell culture.

The invention further concerns antagonists and agonists of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide that inhibit one or more of the functions or activities of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide.

In a further embodiment, the invention concerns isolated nucleic acid molecules that hybridize to the complement of the nucleic acid molecules encoding the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides. The nucleic acid preferably is DNA, and hybridization preferably occurs under stringent conditions. Such nucleic acid molecules can act as antisense molecules of the amplified genes identified herein, which, in turn, can find use in the modulation of the respective amplified genes, or as antisense primers in amplification reactions. Furthermore, such sequences can be used as part of ribozyme and/or triple helix sequence which, in turn, may be used in regulation of the amplified genes.

In another embodiment, the invention concerns a method for determining the presence of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide comprising exposing a cell suspected of containing the polypeptide to an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody and determining binding of the antibody to the cell.

In yet another embodiment, the present invention concerns a method of diagnosing an immune related disease in a mammal, comprising detecting the level of expression of a gene encoding a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide (a) in a test sample of tissue cells obtained from the mammal, and (b) in a control sample of known normal tissue cells of the same cell type, wherein a higher expression level in the test sample indicates the presence of immune related disease in the mammal from which the test tissue cells were obtained.

In another embodiment, the present invention concerns a method of diagnosing an immune disease in a mammal, comprising (a) contacting an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody with a test sample of tissue cells obtained from the mammal, and (b) detecting the formation of a complex between the antibody and the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide in the test sample. The detection may be qualitative or quantitative, and may be performed in comparison with monitoring the complex formation in a control sample of known normal tissue cells of the same cell type. A larger quantity of complexes formed in the test sample indicates

the presence of tumor in the mammal from which the test tissue cells were obtained. The antibody preferably carries a detectable label. Complex formation can be monitored, for example, by light microscopy, flow cytometry, fluorimetry, or other techniques known in the art. The test sample is usually obtained from an individual suspected of having a deficiency or abnormality of the immune system.

5 In another embodiment, the present invention concerns a diagnostic kit, containing an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody and a carrier (e.g. a buffer) in suitable packaging. The kit preferably contains instructions for using the antibody to detect the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide.

In a further embodiment, the invention concerns an article of manufacture, comprising:  
 10 a container;  
 a label on the container; and  
 a composition comprising an active agent contained within the container; wherein the composition is effective for stimulating or inhibiting an immune response in a mammal, the label on the container indicates that the composition can be used to treat an immune related disease, and the active agent in the composition is  
 15 an agent stimulating or inhibiting the expression and/or activity of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. In a preferred aspect, the active agent is a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide or an anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibody.

A further embodiment is a method for identifying a compound capable of inhibiting the expression  
 20 and/or activity of a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide by contacting a candidate compound with a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide under conditions and for a time sufficient to allow these two components to interact. In a specific aspect, either the candidate compound or the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide is immobilized on a solid support. In another aspect, the non-immobilized  
 25 component carries a detectable label.

#### Brief Description of the Drawings

Figure 1 shows the nucleotide sequence (SEQ ID NO:1) of a native sequence PRO245 cDNA, wherein the nucleotide sequence is designated herein as "UNQ219" and/or "DNA35638".

Figure 2 shows the amino acid sequence (SEQ ID NO:2) derived from the nucleotide sequence  
 30 shown in Figure 1.

Figures 3A and 3B show an alignment of nucleotide sequences (SEQ ID NOS:8-11) from a variety of expressed sequence tags as well as a consensus nucleotide sequence derived therefrom designated "DNA30954" (SEQ ID NO:7).

Figure 4 shows a BLAST sequence alignment analysis of a portion of the PRO245 amino acid  
 35 sequence derived from the DNA35638 molecule ("DNA35638") (SEQ ID NO:12) with the human c-mycb ("HSU22376\_2") (SEQ ID NO:3).

Figures 5A, 5B and 5C show the nucleotide sequence comprising a native sequence egf-like homologue cDNA. These are also indicated as SEQ ID NO: 13, SEQ ID NO: 14 and SEQ ID NO: 15, respectively.

Figures 6A, 6B and 6C show the amino acid sequences encoded by the coding sequences of the  
 40 nucleotides described in Figures 5A, 5B and 5C. These polypeptide sequences are also identified as SEQ ID NO: 16, SEQ ID NO: 17 and SEQ ID NO: 18 (PRO217), respectively.

Figures 7A, 7B and 7C show an alignment comparison between prior art sequences used to create DNA28726 (SEQ ID NO: 19), DNA28730 (SEQ ID NO: 20) and DNA28760 (SEQ ID NO: 21), respectively, virtual sequences which were used in the isolation of the nucleotide sequences of the invention. Figure 7A indicates the alignment between Incyte EST sequences 2305118 (SEQ ID NO: 22), 2544914 (SEQ ID NO: 23), 1682522 (SEQ ID NO: 24), 424333 (SEQ ID NO: 25), 640534 (SEQ ID NO: 26), 2211568 (SEQ ID NO: 27), 1436024 (SEQ ID NO: 28), 1600521 (SEQ ID NO: 30), 732577 (SEQ ID NO: 31), 931313 (SEQ ID NO: 33), 045517 (SEQ ID NO: 34), 1557825 (SEQ ID NO: 35), 1555649 (SEQ ID NO: 36), and GenBank sequences W24885 (SEQ ID NO: 29), N95751 (SEQ ID NO: 32). Figure 7B indicates the alignment between Incyte EST sequences 2398238 (SEQ ID NO: 37), 1842628 (SEQ ID NO: 38), 2191592 (SEQ ID NO: 39), 1932631 (SEQ ID NO: 40), 1700782 (SEQ ID NO: 44) and GenBank sequences AA195267 (SEQ ID NO: 41), H99879 (SEQ ID NO: 42), AA195084 (SEQ ID NO: 43). Figure 7C indicates the alignment between GenBank sequences W27896 (SEQ ID NO: 33), W27851 (SEQ ID NO: 46), W22553 (SEQ ID NO: 47), W23268 (SEQ ID NO: 48), W28670 (SEQ ID NO: 50), W27944 (SEQ ID NO: 51), R55894 (SEQ ID NO: 53), W37154 (SEQ ID NO: 57), W38638 (SEQ ID NO: 59) and Incyte EST sequences 400252 (SEQ ID NO: 49), 399998 (SEQ ID NO: 52), 660500 (SEQ ID NO: 54), 662092 (SEQ ID NO: 55), 1682022 (SEQ ID NO: 56), 1577139 (SEQ ID NO: 58).

Figure 8 shows oligonucleotide sequences 28726.p (SEQ ID NO: 60), 28726.f (SEQ ID NO: 61) and 28726.r (SEQ ID NO: 62), which were used in the isolation of DNA32279 (SEQ ID NO: 13), also indicated in Figure 5A.

Figure 9 shows oligonucleotide sequences 28730.p (SEQ ID NO: 63), 28730.f (SEQ ID NO: 64) and 28730.r (SEQ ID NO: 65), which were used in the isolation of DNA32292 (SEQ ID NO: 14), also indicated in Figure 5B.

Figure 10 shows oligonucleotide sequences 28760.p (SEQ ID NO: 66), 28760.f (SEQ ID NO: 67) and 28760.r (SEQ ID NO: 68), which were used in the isolation of DNA33094 (SEQ ID NO: 15), also indicated in Figure 5C.

Figure 11 describes the Blast score, match, percent homology alignment between the coding protein of DNA32279 (SEQ ID NO: 13), a full-length EGF-like homologue of the invention in comparison with GEN12205 (SEQ ID NO: 69), an epidermal growth factor-like protein S1-5.

Figures 12A and 12B describe the Blast score, match and percent homology alignment between the coding protein of DNA32292 (SEQ ID NO: 14), a full-length EGF-like homologue of the invention in comparison with PAC6\_RAT (SEQ ID NO: 70), a serine protease pc6 precursor from rattus norvegicus and FBLC\_MOUSE (SEQ ID NO: 71), a Fibulin-1 isoform c precursor from mus musculus, respectively, each of which contain a cysteine-rich domain which may form EGF-like structures.

Figures 13A and 13B describe the Blast score, match and percent homology alignment between the coding protein of DNA33094 (SEQ ID NO: 15), a full-length EGF-like homologue of the invention in comparison with A43902 (SEQ ID NO: 72), a fragment of eastern newt tanascin, and HSTNX12\_1 (SEQ ID NO: 73), a human tanascin-X precursor, respectively, each of which contain Sistine-rich domains which may form EGF-like structures.

Figure 14 shows the derived amino acid sequence of a native sequence PRO301 polypeptide (SEQ ID NO: 74). This polypeptide is 299 amino acids long, having signal sequence at residue 1 to 27, an extracellular domain at residue 28 to about 258, Ig superfamily homology at residue 94 to 235, a potential transmembrane domain at residue 236 to about 258, and an intracellular domain at about residue 259 to 299.

Figure 15 shows the nucleotide sequence of a native sequence DNA40628 cDNA (SEQ ID NO: 75).

Figure 16 shows the alignment comparison between sequences used to create DNA35936 (SEQ ID NO:76) (from DNA (SEQ ID NOS: 88-91)) from which the consensus sequence used for cloning the cDNA DNA40628 was created.

Figure 17 shows the alignment comparison between DNA35936 (SEQ ID NO:76) (from DNA) and further sequences from the LIFESEQ™ database (Incyte Pharmaceuticals, Palo Alto, CA) and GenBank (SEQ ID NOS:92-235), which were used to extend the from DNA to obtain a consensus sequence shown in the bottom line of the Figure as "consen01" (SEQ ID NO:77).

Figures 18A-18F show the oligonucleotide sequences OLI2162 (35936.f1) (SEQ ID NO:78); OLI2163 (35936.p1) (SEQ ID NO:79); OLI2164 (35936.f2) (SEQ ID NO:80); OLI2165 (35936.r1) (SEQ ID NO:81); OLI2166 (35936.f3) (SEQ ID NO:82); OLI2167 (35936.r2) (SEQ ID NO:83) which were used in the isolation of DNA40628.

Figure 19 describes the Blast score, match and percent homology alignment between 2 overlapping fragments of DNA40628 and A33\_HUMAN, an human A33 antigen precursor. The first fragment compares the coded residues beginning at nucleotide position 121 to 816 of DNA40628 (SEQ ID NO:84) with nucleotides 17 to 284 of A33\_HUMAN (SEQ ID NO:85); The second fragment compares nucleotides 112 to 810 (SEQ ID NO:86) with nucleotides 12 to 284 (SEQ ID NO:87), respectively.

Figures 20A and 20B show a nucleotide sequence (SEQ ID NO:236) containing the nucleotide sequence (SEQ ID NO:237) of a native sequence PRO266 cDNA, wherein the nucleotide sequence (SEQ ID NO:236) is a clone designated herein as "UNQ233" and/or "DNA37150-seq min". Also presented (circled in Figure 20A) is the position of the initiator methionine residue (residues 1-3 of SEQ ID NO: 237; residues 107-109 of SEQ ID NO: 236). The putative transmembrane domain of the protein is encoded by nucleotides beginning at nucleotide 1843 of SEQ ID NO: 237, underlined in Figure 20B. Also in Figure 20B, the stop codon is circled, immediately after the last nucleotide of SEQ ID NO: 237.

Figure 21 shows the amino acid sequence (SEQ ID NO:238) derived from SEQ ID NO:237 shown in Figures 20A and 20B.

Figures 22A-22D show a BLAST sequence alignment analysis of portions of the PRO266 amino acid sequence derived from SEQ ID NO: 237 with portions of the SLIT protein precursor from drosophila melanogaster (SEQ ID NOS:239-247).

Figures 23A-23D show a BLAST sequence alignment analysis of portions of the PRO266 amino acid sequence derived from SEQ ID NO:237 with portions of the Drosophila SLIT protein involved in axon pathway development (SEQ ID NOS. 248-256).

Figure 24 shows an expression sequence tag (SEQ ID NO:257) which was used to form primers herein.

Figures 25A and 25B show the nucleic acid sequence (SEQ ID NO:261) comprising the coding nucleic acid (SEQ ID NO:262) of a native PRO335 polypeptide derived from SEQ ID NO:262. SEQ ID NO:262 begins with at position 65 of SEQ ID NO: 261. The start codon, nucleic acid positions 1-3 of SEQ ID NO:262 is circled. The stop codon is circled, after the last nucleic acid of SEQ ID NO:262, at 3177.

Figure 26 shows the amino acid sequence of PRO335 (SEQ ID NO:263).

Figures 27A and 27B show an alignment of nucleotide sequences from a variety of expressed sequence tags as well as a consensus nucleotide sequence derived therefrom designated "DNA36685", (SEQ ID NO:264) which was used in the process of identifying PRO335, 331, and 326. The expressed sequence tags shown are designated as follows: W22274 (SEQ ID NO:265); and R55603 (SEQ ID NO:266).

Figures 28A through 28C show the results of a BLAST search against PRO335 and amino acid alignments between portions of PRO335 and portions of LIG-1 (SEQ ID NOS:267-269).

Figures 29A and 29B show the amino acid sequence of LIG-1 (SEQ ID NO:270) and the leucine rich repeat domains of LIG-1.

5 Figure 30A through 30C show sequence information related to SEQ ID NO:286 (Figure 30A). Figure 30B shows the results of a BLAST search using SEQ ID NO:286 and Figure 6A shows primers (SEQ ID NOS:287-289) synthesized based on SEQ ID NO:286.

Figure 31 shows primers (SEQ ID NOS:271-278) related to the identification of SEQ ID NO:261.

10 Figure 32 shows the nucleic acid sequence (SEQ ID NO:279) comprising the coding nucleic acid (SEQ ID NO:280) of a native PRO331 polypeptide derived from SEQ ID NO:280. SEQ ID NO:280 begins with the start codon, nucleic acid positions 1-3 of SEQ ID NO:280, circled. The stop codon is also circled, after the last nucleic acid of SEQ ID NO:280, at 1920.

Figure 33 shows the amino acid sequence of PRO331 (SEQ ID NO:281) wherein the signal peptide is shown in parenthesis, and the start of the mature peptide or extracellular domain is shown underlined. The 15 start and end of the leucine rich repeat domains have an X underneath the perspective amino acid. The start of the transmembrane domain is marked with a circle underneath the perspective amino acid. The start of the intracellular domain is marked with a triangle underneath the perspective amino acid.

Figures 34A through 34E show the results of a BLAST search against PRO331 and amino acid alignments between portions of PRO331 and portions of LIG-1 (SEQ ID NOS:282-292).

20 Figures 35A and 35B show the results of a BLAST search (Figure 35A) against SEQ ID NO:264 and amino acid alignments between portions of the amino acid sequence for which SEQ ID NO:4 encodes, (SEQ ID NO:310) and portions of LIG-1 (SEQ ID NOS:293 and 294).

Figure 36 shows primers (SEQ ID NOS:295-297) related to the identification of SEQ ID NO:280.

25 Figures 37A through 37C show the nucleic acid sequence (SEQ ID NO:298) comprising the coding nucleic acid (SEQ ID NO:299) of a native PRO326 polypeptide derived from SEQ ID NO:299. SEQ ID NO:299 begins with the start codon, nucleic acid positions 1-3 of SEQ ID NO:299, circled. The stop codon is also circled, after the last nucleic acid of SEQ ID NO:299, at position 3357.

Figure 38 shows the amino acid sequence of PRO326 (SEQ ID NO:300).

30 Figures 39A through 39D show the results of a BLAST search against PRO326 and amino acid alignments between portions of PRO326 and portions of LIG-1 (SEQ ID NOS:301-303).

Figure 40 shows primers (SEQ ID NOS:304-306) related to the identification of SEQ ID NO:299.

Figure 41 shows additional primers (SEQ ID NOS:307-308) related to the identification of SEQ ID NO:299.

#### Detailed Description of the Preferred Embodiments

##### 35 I. Definitions

The term "immune related disease" means a disease in which a component of the immune system of a mammal causes, mediates or otherwise contributes to a morbidity in the mammal. Also included are diseases in which stimulation or intervention of the immune response has an ameliorative effect on progression of the disease. Included within this term are immune-mediated inflammatory diseases, non-40 immune-mediated inflammatory diseases, infectious diseases, immunodeficiency diseases, neoplasia, etc.

The term "T cell mediated" disease means a disease in which T cells directly or indirectly mediate or otherwise contribute to a morbidity in a mammal. The T cell mediated disease may be associated with cell

mediated effects, lymphokine mediated effects, etc., and even effects associated with B cells if the B cells are stimulated, for example, by the lymphokines secreted by T cells.

Examples of immune-related and inflammatory diseases, some of which are immune or T cell mediated, which can be treated according to the invention include systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic arthritis, spondyloarthropathies, systemic sclerosis (scleroderma), idiopathic inflammatory myopathies (dermatomyositis, polymyositis), Sjögren's syndrome, systemic vasculitis, sarcoidosis, autoimmune hemolytic anemia (immune pancytopenia, paroxysmal nocturnal hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic purpura, immune-mediated thrombocytopenia), thyroiditis (Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic thyroiditis), diabetes mellitus, immune-mediated renal disease (glomerulonephritis, tubulointerstitial nephritis), demyelinating diseases of the central and peripheral nervous systems such as multiple sclerosis, idiopathic demyelinating polyneuropathy or Guillain-Barré syndrome, and chronic inflammatory demyelinating polyneuropathy, hepatobiliary diseases such as infectious hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune chronic active hepatitis, primary biliary cirrhosis, granulomatous hepatitis, and sclerosing cholangitis, inflammatory and fibrotic lung diseases such as inflammatory bowel disease (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and Whipple's disease, autoimmune or immune-mediated skin diseases including bullous skin diseases, erythema multiforme and contact dermatitis, psoriasis, allergic diseases such as asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity and urticaria, immunologic diseases of the lung such as eosinophilic pneumonias, idiopathic pulmonary fibrosis and hypersensitivity pneumonitis, transplantation associated diseases including graft rejection and graft - versus-host-disease. Infectious diseases include AIDS (HIV infection), hepatitis A, B, C, D, and E, bacterial infections, fungal infections, protozoal infections and parasitic infections.

"Treatment" is an intervention performed with the intention of preventing the development or altering the pathology of a disorder. Accordingly, "treatment" refers to both therapeutic treatment and prophylactic or preventative measures. Those in need of treatment include those already with the disorder as well as those in which the disorder is to be prevented. In treatment of an immune related disease, a therapeutic agent may directly decrease or increase the magnitude of response of a component of the immune response, or render the disease more susceptible to treatment by other therapeutic agents, e.g. antibiotics, antifungals, anti-inflammatory agents, chemotherapeutics, etc.

The "pathology" of an immune related disease includes all phenomena that compromise the well-being of the patient. This includes, without limitation, abnormal or uncontrollable cell growth (neutrophilic, eosinophilic, monocytic, lymphocytic cells), antibody production, auto-antibody production, complement production, interference with the normal functioning of neighboring cells, release of cytokines or other secretory products at abnormal levels, suppression or aggravation of any inflammatory or immunological response, infiltration of inflammatory cells (neutrophilic, eosinophilic, monocytic, lymphocytic) into cellular spaces, etc.

"Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, horses, cats, cows, etc. Preferably, the mammal is human.

Administration "in combination with" one or more further therapeutic agents includes simultaneous (concurrent) and consecutive administration in any order.

The term "cytotoxic agent" as used herein refers to a substance that inhibits or prevents the function of cells and/or causes destruction of cells. The term is intended to include radioactive isotopes (e.g.  $^{131}\text{I}$ ,  $^{125}\text{I}$ ,  $^{90}\text{Y}$  and  $^{186}\text{Re}$ ), chemotherapeutic agents, and toxins such as enzymatically active toxins of bacterial, fungal, plant or animal origin, or fragments thereof.

5 A "chemotherapeutic agent" is a chemical compound useful in the treatment of cancer. Examples of chemotherapeutic agents include adriamycin, doxorubicin, epirubicin, 5-fluorouracil, cytosine arabinoside ("Ara-C"), cyclophosphamide, thiotepa, busulfan, cytoxin, taxoids, e.g. paclitaxel (Taxol, Bristol-Myers Squibb Oncology, Princeton, NJ), and doxetaxel (Taxotere, Rh $\square$ ne-Poulenc Rorer, Antony, Rnace), toxotere, methotrexate, cisplatin, melphalan, vinblastine, bleomycin, etoposide, ifosfamide, mitomycin C, 10 mitoxantrone, vincristine, vinorelbine, carboplatin, teniposide, daunomycin, carminomycin, aminopterin, dactinomycin, mitomycins, esperamicins (see U.S. Pat. No. 4,675,187), melphalan and other related nitrogen mustards. Also included in this definition are hormonal agents that act to regulate or inhibit hormone action on tumors such as tamoxifen and onapristone.

A "growth inhibitory agent" when used herein refers to a compound or composition which inhibits 15 growth of a cell, especially cancer cell overexpressing any of the genes identified herein, either *in vitro* or *in vivo*. Thus, the growth inhibitory agent is one which significantly reduces the percentage of cells overexpressing such genes in S phase. Examples of growth inhibitory agents include agents that block cell cycle progression (at a place other than S phase), such as agents that induce G1 arrest and M-phase arrest. Classical M-phase blockers include the vincas (vincristine and vinblastine), taxol, and topo II inhibitors such 20 as doxorubicin, epirubicin, daunorubicin, etoposide, and bleomycin. Those agents that arrest G1 also spill over into S-phase arrest, for example, DNA alkylating agents such as tamoxifen, prednisone, dacarbazine, mechlorethamine, cisplatin, methotrexate, 5-fluorouracil, and ara-C. Further information can be found in *The Molecular Basis of Cancer*, Mendelsohn and Israel, eds., Chapter 1, entitled "Cell cycle regulation, oncogens, and antineoplastic drugs" by Murakami *et al.* (WB Saunders: Philadelphia, 1995), especially p. 13.

25 The term "cytokine" is a generic term for proteins released by one cell population which act on another cell as intercellular mediators. Examples of such cytokines are lymphokines, monokines, and traditional polypeptide hormones. Included among the cytokines are growth hormone such as human growth hormone, N-methionyl human growth hormone, and bovine growth hormone; parathyroid hormone; thyroxine; insulin; proinsulin; relaxin; prorelaxin; glycoprotein hormones such as follicle stimulating hormone 30 (FSH), thyroid stimulating hormone (TSH), and luteinizing hormone (LH); hepatic growth factor; fibroblast growth factor; prolactin; placental lactogen; tumor necrosis factor- $\alpha$  and - $\beta$ ; mullerian-inhibiting substance; mouse gonadotropin-associated peptide; inhibin; activin; vascular endothelial growth factor; integrin; thrombopoietin (TPO); nerve growth factors such as NGF- $\beta$ ; platelet-growth factor; transforming growth factors (TGFs) such as TGF- $\alpha$  and TGF- $\beta$ ; insulin-like growth factor-I and -II; erythropoietin (EPO); 35 osteoinductive factors; interferons such as interferon- $\alpha$ , - $\beta$ , and - $\gamma$ ; colony stimulating factors (CSFs) such as macrophage-CSF (M-CSF); granulocyte-macrophage-CSF (GM-CSF); and granulocyte-CSF (G-CSF); interleukins (ILs) such as IL-1, IL-1 $\alpha$ , IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-11, IL-12; a tumor necrosis factor such as TNF- $\alpha$  or TNF- $\beta$ ; and other polypeptide factors including LIF and kit ligand (KL). As used herein, the term cytokine includes proteins from natural sources or from recombinant cell culture and 40 biologically active equivalents of the native sequence cytokines.

As used herein, a "PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide" refers to a native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 having the same amino acid sequence as a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 derived from nature. Such native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be isolated from nature or can be produced by recombinant and/or synthetic means. The term specifically encompasses naturally-occurring truncated or secreted forms (e.g., an extracellular domain sequence), naturally-occurring variant forms (e.g., alternatively spliced forms) and naturally-occurring allelic variants of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326. In one embodiment of the invention, the native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 is a mature or full-length native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 comprising amino acids 1-312 of Figure 2 (SEQ ID NO:2), 1-379 of Figure 6C (SEQ ID NO:18), 1-299 of Figure 14 (SEQ ID NO:74), 1-696 of Figure 21 (SEQ ID NO:238), 1-1059 of Figure 26 (SEQ ID NO:263), 1-640 of Figure 33 (SEQ ID NO:281) or 1-1119 of Figure 38 (SEQ ID NO:300).

The term "polypeptide of the invention" refers to each individual PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. All disclosures in this specification which refer to the "polypeptide of the invention" or to "the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide" refer to each of the polypeptides individually as well as jointly. For example, descriptions of the preparation of, purification of, derivation of, formation of antibodies to or against, administration of, compositions containing, treatment of a disease with, etc., pertain to each polypeptide of the invention individually. The term "compound of the invention" includes the polypeptide of the invention, as well as agonist antibodies for and antagonist antibodies to these polypeptide, peptides or small molecules having agonist or antagonist activity developed from the polypeptide, etc.

An "isolated" nucleic acid molecule encoding a polypeptide of the invention is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the polypeptide-encoding nucleic acid. An isolated polypeptide-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated nucleic acid molecules therefore are distinguished from the polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated nucleic acid molecule encoding a polypeptide of the invention includes polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express a polypeptide of the invention where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is



accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency of hybridization reactions, see Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 ug/ml), 0.1% SDS, and 10% dextran sulfate at 42C, with washes at 42C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55C.

"Moderately stringent conditions" may be identified as described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing solution and hybridization conditions (e.g., temperature, ionic strength and %SDS) less stringent than those described above. An example of moderately stringent conditions is overnight incubation at 37C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/mL denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

The term "epitope tagged" when used herein refers to a chimeric polypeptide comprising a polypeptide of the invention fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody can be made, yet is short enough such that it does not interfere with activity of the polypeptide to which it is fused. The tag polypeptide preferably also is fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid residues and usually between about 8 and 50 amino acid residues (preferably, between about 10 and 20 amino acid residues).

"Active" or "activity" in the context of variants of the polypeptide of the invention refers to form(s) of proteins of the invention which retain the biologic and/or immunologic activities of a native or naturally-occurring polypeptide of the invention.

"Biological activity" in the context of an antibody or another molecule that can be identified by the screening assays disclosed herein (e.g. an organic or inorganic small molecule, peptide, etc.) is used to refer to

the ability of such molecules to induce or inhibit infiltration of inflammatory cells into a tissue, to stimulate or inhibit T-cell proliferation and to stimulate or inhibit lymphokine release by cells. Another preferred activity is increased vascular permeability or the inhibition thereof.

The term "antagonist" is used in the broadest sense, and includes any molecule that partially or fully blocks, inhibits, or neutralizes a biological activity of a native polypeptide of the invention disclosed herein. In a similar manner, the term "agonist" is used in the broadest sense and includes any molecule that mimics a biological activity of a native polypeptide of the invention disclosed herein. Suitable agonist or antagonist molecules specifically include agonist or antagonist antibodies or antibody fragments, fragments or amino acid sequence variants of native polypeptides of the invention, peptides, small organic molecules, etc.

10 A "small molecule" is defined herein to have a molecular weight below about 600 daltons.

"Antibodies" (Abs) and "immunoglobulins" (Igs) are glycoproteins having the same structural characteristics. While antibodies exhibit binding specificity to a specific antigen, immunoglobulins include both antibodies and other antibody-like molecules which lack antigen specificity. Polypeptides of the latter kind are, for example, produced at low levels by the lymph system and at increased levels by myelomas. The term "antibody" is used in the broadest sense and specifically covers, without limitation, intact monoclonal antibodies, polyclonal antibodies, multispecific antibodies (e.g. bispecific antibodies) formed from at least two intact antibodies, and antibody fragments so long as they exhibit the desired biological activity.

"Native antibodies" and "native immunoglobulins" are usually heterotetrameric glycoproteins of about 150,000 daltons, composed of two identical light (L) chains and two identical heavy (H) chains. Each light chain is linked to a heavy chain by one covalent disulfide bond, while the number of disulfide linkages varies among the heavy chains of different immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulfide bridges. Each heavy chain has at one end a variable domain ( $V_H$ ) followed by a number of constant domains. Each light chain has a variable domain at one end ( $V_L$ ) and a constant domain at its other end; the constant domain of the light chain is aligned with the first constant domain of the heavy chain, and the light-chain variable domain is aligned with the variable domain of the heavy chain. Particular amino acid residues are believed to form an interface between the light- and heavy-chain variable domains.

The term "variable" refers to the fact that certain portions of the variable domains differ extensively in sequence among antibodies and are used in the binding and specificity of each particular antibody for its particular antigen. However, the variability is not evenly distributed throughout the variable domains of antibodies. It is concentrated in three segments called complementarity-determining regions (CDRs) or hypervariable regions both in the light-chain and the heavy-chain variable domains. The more highly conserved portions of variable domains are called the framework (FR). The variable domains of native heavy and light chains each comprise four FR regions, largely adopting a beta-sheet configuration, connected by three CDRs, which form loops connecting, and in some cases forming part of, the beta-sheet structure. The CDRs in each chain are held together in close proximity by the FR regions and, with the CDRs from the other chain, contribute to the formation of the antigen-binding site of antibodies (see Kabat *et al.*, *NIH Publ. No. 91-3242*, Vol. I, pages 647-669 (1991)). The constant domains are not involved directly in binding an antibody to an antigen, but exhibit various effector functions, such as participation of the antibody in antibody-dependent cellular toxicity.

"Antibody fragments" comprise a portion of an intact antibody, preferably the antigen binding or variable region of the intact antibody. Examples of antibody fragments include Fab, Fab', F(ab')<sub>2</sub>, and Fv

fragments: diabodies; linear antibodies (Zapata *et al.*, Protein Eng. 8(10):1057-1062 [1995]); single-chain antibody molecules; and multispecific antibodies formed from antibody fragments.

Papain digestion of antibodies produces two identical antigen-binding fragments, called "Fab" fragments, each with a single antigen-binding site, and a residual "Fc" fragment, whose PRO245, PRO217, PRO301, Pro266, pro335, pro331 or pro326 reflects its ability to crystallize readily. Pepsin treatment yields an F(ab')<sub>2</sub> fragment that has two antigen-combining sites and is still capable of cross-linking antigen.

"Fv" is the minimum antibody fragment which contains a complete antigen-recognition and -binding site. This region consists of a dimer of one heavy- and one light-chain variable domain in tight, non-covalent association. It is in this configuration that the three CDRs of each variable domain interact to define an antigen-binding site on the surface of the V<sub>H</sub>-V<sub>L</sub> dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

The Fab fragment also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. F(ab')<sub>2</sub> antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

The "light chains" of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa (κ) and lambda (λ), based on the amino acid sequences of their constant domains.

Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), e.g., IgG1, IgG2, IgG3, IgG4, IgA, and IgA2. The heavy-chain constant domains that correspond to the different classes of immunoglobulins are called α, δ, ε, γ, and μ, respectively. The subunit structures and three-dimensional configurations of different classes of immunoglobulins are well known.

The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. In addition to their specificity, the monoclonal antibodies are advantageous in that they are synthesized by the hybridoma culture, uncontaminated by other immunoglobulins. The modifier "monoclonal" indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by Kohler *et al.*, Nature, 256:495 [1975], or may be made by recombinant DNA methods (see, e.g., U.S. Patent No. 4,816,567). The "monoclonal antibodies" may also be

isolated from phage antibody libraries using the techniques described in Clackson *et al.*, Nature, 352:624-628 [1991] and Marks *et al.*, J. Mol. Biol., 222:581-597 (1991), for example. See also U.S. Patent Nos. 5,750,373, 5,571,698, 5,403,484 and 5,223,409 which describe the preparation of antibodies using phagemid and phage vectors.

5 The monoclonal antibodies herein specifically include "chimeric" antibodies (immunoglobulins) in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such  
10 antibodies, so long as they exhibit the desired biological activity (U.S. Patent No. 4,816,567; Morrison *et al.*, Proc. Natl. Acad. Sci. USA, 81:6851-6855 [1984]).

"Humanized" forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. For  
15 the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a complementarity-determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity, and capacity. In some instances, Fv framework region (FR) residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibodies may comprise residues which are  
20 found neither in the recipient antibody nor in the imported CDR or framework sequences. These modifications are made to further refine and maximize antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody  
25 optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. For further details, see Jones *et al.*, Nature, 321:522-525 (1986); Reichmann *et al.*, Nature, 332:323-329 [1988]; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992). The humanized antibody includes a "primatized" antibody where the antigen-binding region of the antibody is derived from an antibody produced by immunizing macaque monkeys with the antigen of interest. Antibodies containing residues from  
30 Old World monkeys are also possible within the invention. See, for example, U.S. Patent Nos. 5,658,570; 5,693,780; 5,681,722; 5,750,105; and 5,756,096.

"Single-chain Fv" or "sFv" antibody fragments comprise the V<sub>H</sub> and V<sub>L</sub> domains of antibody, wherein these domains are present in a single polypeptide chain. Preferably, the Fv polypeptide further comprises a polypeptide linker between the V<sub>H</sub> and V<sub>L</sub> domains which enables the sFv to form the desired  
35 structure for antigen binding. For a review of sFv see Pluckthun in *The Pharmacology of Monoclonal Antibodies*, vol. 113, Rosenberg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994).

The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy-chain variable domain (V<sub>H</sub>) connected to a light-chain variable domain (V<sub>L</sub>) in the same polypeptide chain (V<sub>H</sub> - V<sub>L</sub>). By using a linker that is too short to allow pairing between the two  
40 domains on the same chain, the domains are forced to pair with the complementary domains of another chain

and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and Hollinger *et al.*, Proc. Natl. Acad. Sci. USA, 90:6444-6448 (1993).

An "isolated" antibody is one which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the compound of the invention will be purified (1) to greater than 95% by weight of the compound as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. Isolated compound, e.g. antibody or polypeptide, includes the compound *in situ* within recombinant cells since at least one component of the compound's natural environment will not be present. Ordinarily, however, isolated compound will be prepared by at least one purification step.

The word "label" when used herein refers to a detectable compound or composition which is conjugated directly or indirectly to the compound, e.g. antibody or polypeptide, so as to generate a "labelled" compound. The label may be detectable by itself (e.g. radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, may catalyze chemical alteration of a substrate compound or composition which is detectable.

By "solid phase" is meant a non-aqueous matrix to which the compound of the present invention can adhere. Examples of solid phases encompassed herein include those formed partially or entirely of glass (e.g., controlled pore glass), polysaccharides (e.g., agarose), polyacrylamides, polystyrene, polyvinyl alcohol and silicones. In certain embodiments, depending on the context, the solid phase can comprise the well of an assay plate; in others it is a purification column (e.g., an affinity chromatography column). This term also includes a discontinuous solid phase of discrete particles, such as those described in U.S. Patent No. 4,275,149.

A "liposome" is a small vesicle composed of various types of lipids, phospholipids and/or surfactant which is useful for delivery of a drug (such as the anti-ErbB2 antibodies disclosed herein and, optionally, a chemotherapeutic agent) to a mammal. The components of the liposome are commonly arranged in a bilayer formation, similar to the lipid arrangement of biological membranes.

As used herein, the term "immunoadhesin" designates antibody-like molecules which combine the binding specificity of a heterologous protein (an "adhesin") with the effector functions of immunoglobulin constant domains. Structurally, the immunoadhesins comprise a fusion of an amino acid sequence with the desired binding specificity which is other than the antigen recognition and binding site of an antibody (i.e., is "heterologous"), and an immunoglobulin constant domain sequence. The adhesin part of an immunoadhesin molecule typically is a contiguous amino acid sequence comprising at least the binding site of a receptor or a ligand. The immunoglobulin constant domain sequence in the immunoadhesin may be obtained from any immunoglobulin, such as IgG-1, IgG-2, IgG-3, or IgG-4 subtypes, IgA (including IgA-1 and IgA-2), IgE, IgD or IgM.

## II. Compositions and Methods of the Invention

### 1. Preparation of the polypeptides of the invention

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO245, PRO217, PRO301, PRO266, PRO335,

PRO331 or PRO326 (UNQ219, UNQ191, UNQ264, UNQ233, UNQ287V, UNQ292 or UNQ287 respectively). In particular, cDNA encoding a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide has been identified and isolated, as disclosed in further detail in the Examples below. It is noted that proteins produced in separate expression rounds may be given different PRO numbers but the UNQ number is unique for any given DNA and the encoded protein, and will not be changed. However, for sake of simplicity, in the present specification the protein encoded by DNA35638, DNA33094, DNA40628, DNA37150, DNA41388, DNA40981 AND DNA37140 as well as all further native homologues and variants included in the foregoing definition of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326, will be referred to as PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 or simply as "the polypeptide of the invention", regardless of their origin or mode of preparation.

The description below relates primarily to production of the polypeptide of the invention by culturing cells transformed or transfected with a vector containing nucleic acid which encodes of the polypeptide of the invention. It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare of the polypeptide of the invention. For instance, the polypeptide sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of the polypeptide of the invention may be chemically synthesized separately and combined using chemical or enzymatic methods to produce the full-length polypeptide.

i. Isolation of DNA Encoding the Polypeptide of the Invention

DNA encoding the polypeptide of the invention may be obtained from a cDNA library prepared from tissue believed to possess the polypeptide mRNA and to express it at a detectable level. Accordingly, human DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The gene encoding the polypeptide of the invention may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to the polypeptide of the invention or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding the polypeptide of the invention is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like <sup>32</sup>P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases.

Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNASTar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or  
5 genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

ii. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for  
10 production of the polypeptides of the invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical  
15 Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example,  $\text{CaPO}_4$  and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall  
20 barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically  
25 carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, e.g., polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

30 Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

35 In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for vectors encoding the polypeptides of the invention. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated polypeptides of the invention are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and  
40 Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40

(COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen. Virol., 36:59 (1977)); Chinese hamster ovary cells/DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

iii. Selection and Use of a Replicable Vector

The nucleic acid (e.g., cDNA or genomic DNA) encoding the polypeptides of the invention may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, phagemid or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The polypeptide of the invention may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the DNA encoding the polypeptide of the invention that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, e.g., the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces*  $\alpha$ -factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2u plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the nucleic acid encoding the polypeptide of the invention, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp1* gene present in the yeast plasmid



YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp1* gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the nucleic acid  
 5 sequence encoding the polypeptide of the invention to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the  $\beta$ -lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (*trp*) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the *tac* promoter [deBoer et al., Proc. Natl. Acad.  
 10 Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the polypeptide of the invention.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase,  
 15 glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2,  
 20 isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

Transcription of the polypeptide of the invention from vectors in mammalian host cells is controlled,  
 25 for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

30 Transcription of a DNA encoding the polypeptide of the invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin,  $\alpha$ -fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40  
 35 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the coding sequence of the polypeptide of the invention, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or  
 40 nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and,

occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding the polypeptide of the invention.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of the polypeptide of the invention in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

iv. Detecting Gene Expression

Gene expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence of the inventive polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to DNA encoding the polypeptide of the invention and encoding a specific antibody epitope.

iii. Purification of Polypeptide

Forms of the polypeptide of the invention may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (e.g. Triton-X 100) or by enzymatic cleavage. Cells employed in expression of the polypeptide of the invention can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify the polypeptide of the invention from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the polypeptide of the invention. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification: Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular polypeptide of the invention produced.

2. Tissue Distribution

The location of tissues expressing the polypeptides of the invention can be identified by determining mRNA expression in various human tissues. The location of such genes provides information about which tissues are most likely to be affected by the stimulating and inhibiting activities of the polypeptides of the

invention. The location of a gene in a specific tissue also provides sample tissue for the activity blocking assays discussed below.

As noted before, gene expression in various tissues may be measured by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA (Thomas, *Proc. Natl. Acad. Sci. USA*, 5 77:5201-5205 [1980]), dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes.

Gene expression in various tissues, alternatively, may be measured by immunological methods, such 10 as immunohistochemical staining of tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence of a polypeptide of the invention or against a synthetic peptide based on the DNA sequences encoding the polypeptide of the invention or against an 15 exogenous sequence fused to a DNA encoding a polypeptide of the invention and encoding a specific antibody epitope. General techniques for generating antibodies, and special protocols for Northern blotting and *in situ* hybridization are provided below.

### 3. Antibody Binding Studies

The activity of the polypeptides of the invention can be further verified by antibody binding studies, 20 in which the ability of anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibodies to inhibit the effect of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptides on tissue cells is tested. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies, the preparation of which will be described hereinbelow.

Antibody binding studies may be carried out in any known assay method, such as competitive 25 binding assays, direct and indirect sandwich assays, and immunoprecipitation assays. Zola, *Monoclonal Antibodies: A Manual of Techniques*, pp.147-158 (CRC Press, Inc., 1987).

Competitive binding assays rely on the ability of a labeled standard to compete with the test sample analyte for binding with a limited amount of antibody. The amount of target protein in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate 30 determining the amount of standard that becomes bound, the antibodies preferably are insolubilized before or after the competition, so that the standard and analyte that are bound to the antibodies may conveniently be separated from the standard and analyte which remain unbound.

Sandwich assays involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, of the protein to be detected. In a sandwich assay, the test sample analyte is 35 bound by a first antibody which is immobilized on a solid support, and thereafter a second antibody binds to the analyte, thus forming an insoluble three-part complex. See, e.g., US Pat No. 4,376,110. The second antibody may itself be labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assay). For example, one type of sandwich assay is an ELISA assay, in which case the detectable moiety is an enzyme.

40 For immunohistochemistry, the tissue sample may be fresh or frozen or may be embedded in paraffin and fixed with a preservative such as formalin, for example.

#### 4. Cell-Based Assays

Cell-based assays and animal models for immune related diseases can be used to further understand the relationship between the genes and polypeptides identified herein and the development and pathogenesis of immune related disease.

- 5 In a different approach, cells of a cell type known to be involved in a particular immune related disease are transfected with the cDNAs described herein, and the ability of these cDNAs to stimulate or inhibit immune function is analyzed. Suitable cells can be transfected with the desired gene, and monitored for immune function activity. Such transfected cell lines can then be used to test the ability of poly- or monoclonal antibodies or antibody compositions to inhibit or stimulate immune function, for example to  
 10 modulate T-cell proliferation or inflammatory cell infiltration. Cells transfected with the coding sequences of the genes identified herein can further be used to identify drug candidates for the treatment of immune related diseases.

In addition, primary cultures derived from transgenic animals (as described below) can be used in the cell-based assays herein, although stable cell lines are preferred. Techniques to derive continuous cell lines  
 15 from transgenic animals are well known in the art (see, e.g. Small *et al.*, Mol. Cell. Biol. 5, 642-648 [1985]).

One suitable cell based assay is the mixed lymphocyte reaction (MLR). Current Protocols in Immunology, unit 3.12; edited by J E Coligan, A M Kruisbeek, D H Marglies, E M Shevach, W Strober, National Institutes of Health, Published by John Wiley & Sons, Inc. In this assay, the ability of a test compound to stimulate the proliferation of activated T cells is assayed. A suspension of responder T cells is  
 20 cultured with allogeneic stimulator cells and the proliferation of T cells is measured by uptake of tritiated thymidine. This assay is a general measure of T cell reactivity. Since the majority of T cells respond to and produce IL-2 upon activation, differences in responsiveness in this assay in part reflect differences in IL-2 production by the responding cells. The MLR results can be verified by a standard lymphokine (IL-2) detection assay. Current Protocols in Immunology, above, 3.15, 6.3.

- 25 A proliferative T cell response in an MLR assay may be due to a mitogenic response or may be due to a stimulatory response by the T cells. Additional verification of the T cell stimulatory activity of the polypeptides of the invention can be obtained by a costimulation assay. T cell activation requires an antigen specific signal mediated through the major histocompatibility complex (MHC) and a costimulatory signal mediated through a second ligand binding interaction, for example, the B7(CD80, CD86)/CD28 binding  
 30 interaction. CD28 crosslinking increases lymphokine secretion by activated T cells. T cell activation has both negative and positive controls through the binding of ligands which have a negative or positive effect. CD28 and CTLA-4 are related glycoproteins in the Ig superfamily which bind to B7. CD28 binding to B7 has a positive costimulation effect of T cell activation; conversely, CTLA-4 binding to B7 has a negative T cell deactivating effect. Chambers, C. A. and Allison, J. P., *Curr. Opin. Immunol.* (1997) 9:396. Schwartz, R. H.,  
 35 *Cell* (1992) 71:1065; Linsey, P. S. and Ledbetter, J. A., *Annu. Rev. Immunol.* (1993) 11:191; June, C. H. et al, *Immunol. Today* (1994) 15:321; Jenkins, M. K., *Immunity* (1994) 1:405. In a costimulation assay, the polypeptides of the invention are assayed for T cell costimulatory or inhibitory activity.

Polypeptides of the invention, as well as other compounds of the invention, which are stimulators (costimulators) of T cell proliferation, as determined by MLR and costimulation assays, for example, are  
 40 useful in treating immune related diseases characterized by poor, suboptimal or inadequate immune function. These diseases are treated by stimulating the proliferation and activation of T cells (and T cell mediated immunity) and enhancing the immune response in a mammal through administration of a stimulatory

compound, such as the stimulating polypeptides of the invention. The stimulating polypeptide may be a PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide or an agonist antibody therefor. Immunoadjuvant therapy for treatment of tumors, described in more detail below, is an example of this use of the stimulating compounds of the invention. Antibodies which bind to inhibitory polypeptides function to enhance the immune response by removing the inhibitory effect of the inhibiting polypeptides. This effect is seen in experiments using anti-CTLA-4 antibodies which enhance T cell proliferation, presumably by removal of the inhibitory signal caused by CTLA-4 binding. Walunas, T. L. et al, *Immunity* (1994) 1:405. This use is also validated in experiments with 4-1BB glycoprotein, a member of the tumor necrosis factor receptor family which binds to a ligand (4-1BBL) expressed on primed T cells and signals T cell activation and growth. Alderson, M. E. et al., *J. Immunol.* (1994) 24:2219. Inhibition of 4-1BB binding by treatment with an anti-4-1BB antibody increases the severity of graft-versus-host disease and may be used to eradicate tumors. Hellstrom, I. and Hellstrom, K. E., *Crit. Rev. Immunol.* (1998) 18:1.

On the other hand, polypeptides of the invention, as well as other compounds of the invention, which are inhibitors of T cell proliferation/activation and/or lymphokine secretion, can be directly used to suppress the immune response. These compounds are useful to reduce the degree of the immune response and to treat immune related diseases characterized by a hyperactive, superoptimal, or autoimmune response. Alternatively, antibodies which bind to the stimulating polypeptides of the invention and block the stimulating effect of these molecules can be used to suppress the T cell mediated immune response by inhibiting T cell proliferation/activation and/or lymphokine secretion. Blocking the stimulating effect of the polypeptides suppresses the immune response of the mammal.

#### 5. Animal Models

The results of the cell based *in vitro* assays can be further verified using *in vivo* animal models and assays for T-cell function. A variety of well known animal models can be used to further understand the role of the genes identified herein in the development and pathogenesis of immune related disease, and to test the efficacy of candidate therapeutic agents, including antibodies, and other antagonists of the native polypeptides, including small molecule antagonists. The *in vivo* nature of such models makes them particularly predictive of responses in human patients. Animal models of immune related diseases include both non-recombinant and recombinant (transgenic) animals. Non-recombinant animal models include, for example, rodent, e.g., murine models. Such models can be generated by introducing cells into syngeneic mice using standard techniques, e.g. subcutaneous injection, tail vein injection, spleen implantation, intraperitoneal implantation, implantation under the renal capsule, etc.

Contact hypersensitivity is a simple *in vivo* assay of cell mediated immune function. In this procedure, epidermal cells are exposed to exogenous haptens which give rise to a delayed type hypersensitivity reaction which is measured and quantitated. Contact sensitivity involves an initial sensitizing phase followed by an elicitation phase. The elicitation phase occurs when the epidermal cells encounter an antigen to which they have had previous contact. Swelling and inflammation occur, making this an excellent model of human allergic contact dermatitis. A suitable procedure is described in detail in *Current Protocols in Immunology*, Eds. J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach and W. Strober, John Wiley & Sons, Inc., 1994, unit 4.2. See also Grabbe, S. and Schwarz, T., *Immun. Today* 19(1):37-44 (1998).

Graft-versus-host disease occurs when immunocompetent cells are transplanted into immunosuppressed or tolerant patients. The donor cells recognize and respond to host antigens. The response can vary from life threatening severe inflammation to mild cases of diarrhea and weight loss. Graft-

versus-host disease models provide a means of assessing T cell reactivity against MHC antigens and minor transplant antigens. A suitable procedure is described in detail in Current Protocols in Immunology, above, unit 4.3.

An animal model for skin allograft rejection is a means of testing the ability of T cells to mediate in vivo tissue destruction which is indicative of and a measure of their role in anti-viral and tumor immunity. The most common and accepted models use murine tail-skin grafts. Repeated experiments have shown that skin allograft rejection is mediated by T cells, helper T cells and killer-effector T cells, and not antibodies. Auchincloss, H. Jr. and Sachs, D. H., Fundamental Immunology, 2nd ed., W. E. Paul ed., Raven Press, NY, 1989, 889-992. A suitable procedure is described in detail in Current Protocols in Immunology, above, unit 4.4. Other transplant rejection models which can be used to test the compounds of the invention are the allogeneic heart transplant models described by Tanabe, M. et al, Transplantation (1994) 58:23 and Tinubu, S. A. et al, J. Immunol. (1994) 4330-4338.

Animal models for delayed type hypersensitivity provides an assay of cell mediated immune function as well. Delayed type hypersensitivity reactions are a T cell mediated in vivo immune response characterized by inflammation which does not reach a peak until after a period of time has elapsed after challenge with an antigen. These reactions also occur in tissue specific autoimmune diseases such as multiple sclerosis (MS) and experimental autoimmune encephalomyelitis (EAE, a model for MS). A suitable procedure is described in detail in Current Protocols in Immunology, above, unit 4.5.

EAE is a T cell mediated autoimmune disease characterized by T cell and mononuclear cell inflammation and subsequent demyelination of axons in the central nervous system. EAE is generally considered to be a relevant animal model for MS in humans. Bolton, C., Multiple Sclerosis (1995) 1:143. Both acute and relapsing-remitting models have been developed. The compounds of the invention can be tested for T cell stimulatory or inhibitory activity against immune mediated demyelinating disease using the protocol described in Current Protocols in Immunology, above, units 15.1 and 15.2. See also the models for myelin disease in which oligodendrocytes or Schwann cells are grafted into the central nervous system as described in Duncan, I. D. et al, Molec. Med. Today (1997) 554-561.

An animal model for arthritis is collagen-induced arthritis. This model shares clinical, histological and immunological characteristics of human autoimmune rheumatoid arthritis and is an acceptable model for human autoimmune arthritis. Mouse and rat models are characterized by synovitis, erosion of cartilage and subchondral bone. The compounds of the invention can be tested for activity against autoimmune arthritis using the protocols described in Current Protocols in Immunology, above, units 15.5. See also the model using a monoclonal antibody to CD18 and VLA-4 integrins described in Issekutz, A. C. et al., Immunology (1996) 88:569.

A model of asthma has been described in which antigen-induced airway hyper-reactivity, pulmonary eosinophilia and inflammation are induced by sensitizing an animal with ovalbumin and then challenging the animal with the same protein delivered by aerosol. Several animal models (guinea pig, rat, non-human primate) show symptoms similar to atopic asthma in humans upon challenge with aerosol antigens. Murine models have many of the features of human asthma. Suitable procedures to test the compounds of the invention for activity and effectiveness in the treatment of asthma are described by Wolyniec, W. W. et al, Am. J. Respir. Cell Mol. Biol. (1998) 18:777 and the references cited therein.

Additionally, the compounds of the invention can be tested on animal models for psoriasis like diseases. Evidence suggests a T cell pathogenesis for psoriasis. The compounds of the invention can be

tested in the scid/scid mouse model described by Schon, M. P. et al, Nat. Med. (1997) 3:183, in which the mice demonstrate histopathologic skin lesions resembling psoriasis. Another suitable model is the human skin/scid mouse chimera prepared as described by Nickoloff, B. J. et al, Am. J. Path. (1995) 146:580.

Recombinant (transgenic) animal models can be engineered by introducing the coding portion of the  
5 genes identified herein into the genome of animals of interest, using standard techniques for producing transgenic animals. Animals that can serve as a target for transgenic manipulation include, without limitation, mice, rats, rabbits, guinea pigs, sheep, goats, pigs, and non-human primates, e.g. baboons, chimpanzees and monkeys. Techniques known in the art to introduce a transgene into such animals include pronucleic microinjection (Hoppe and Wanger, U.S. Patent No. 4,873,191); retrovirus-mediated gene transfer  
10 into germ lines (e.g., Van der Putten *et al.*, Proc. Natl. Acad. Sci. USA 82, 6148-615 [1985]); gene targeting in embryonic stem cells (Thompson *et al.*, Cell 56, 313-321 [1989]); electroporation of embryos (Lo, Mol. Cel. Biol. 3, 1803-1814 [1983]); sperm-mediated gene transfer (Lavitrano *et al.*, Cell 57, 717-73 [1989]). For review, see, for example, U.S. Patent No. 4,736,866.

For the purpose of the present invention, transgenic animals include those that carry the transgene  
15 only in part of their cells ("mosaic animals"). The transgene can be integrated either as a single transgene, or in concatamers, e.g., head-to-head or head-to-tail tandems. Selective introduction of a transgene into a particular cell type is also possible by following, for example, the technique of Lasko *et al.*, Proc. Natl. Acad. Sci. USA 89, 6232-636 (1992).

The expression of the transgene in transgenic animals can be monitored by standard techniques. For  
20 example, Southern blot analysis or PCR amplification can be used to verify the integration of the transgene. The level of mRNA expression can then be analyzed using techniques such as *in situ* hybridization, Northern blot analysis, PCR, or immunocytochemistry.

The animals may be further examined for signs of immune disease pathology, for example by histological examination to determine infiltration of immune cells into specific tissues. Blocking experiments  
25 can also be performed in which the transgenic animals are treated with the compounds of the invention to determine the extent of the T cell proliferation stimulation or inhibition of the compounds. In these experiments, blocking antibodies which bind to the polypeptide of the invention, prepared as described above, are administered to the animal and the effect on immune function is determined.

Alternatively, "knock out" animals can be constructed which have a defective or altered gene  
30 encoding a polypeptide identified herein, as a result of homologous recombination between the endogenous gene encoding the polypeptide and altered genomic DNA encoding the same polypeptide introduced into an embryonic cell of the animal. For example, cDNA encoding a particular polypeptide can be used to clone genomic DNA encoding that polypeptide in accordance with established techniques. A portion of the genomic DNA encoding a particular polypeptide can be deleted or replaced with another gene, such as a gene  
35 encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected  
40 cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant

female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the polypeptide.

#### 6. ImmunoAdjuvant Therapy

In one embodiment, the immunostimulating compounds of the invention can be used in immunoadjuvant therapy for the treatment of tumors (cancer). It is now well established that T cells recognize human tumor specific antigens. One group of tumor antigens, encoded by the MAGE, BAGE and GAGE families of genes, are silent in all adult normal tissues, but are expressed in significant amounts in tumors, such as melanomas, lung tumors, head and neck tumors, and bladder carcinomas. DeSmet, C. et al, (1996) Proc. Natl. Acad. Sci. USA, 93:7149. It has been shown that costimulation of T cells induces tumor regression and an antitumor response both in vitro and in vivo. Melero, I. et al, Nature Medicine (1997) 3:682; Kwon, E. D. et al, Proc. Natl. Acad. Sci. USA (1997) 94:8099; Lynch, D. H. et al, Nature Medicine (1997) 3:625; Finn, O. J. and Lotze, M. T., J. Immunol. (1998) 21:114. The stimulatory compounds of the invention can be administered as adjuvants, alone or together with a growth regulating agent, cytotoxic agent or chemotherapeutic agent, to stimulate T cell proliferation/activation and an antitumor response to tumor antigens. The growth regulating, cytotoxic, or chemotherapeutic agent may be administered in conventional amounts using known administration regimes. Immunostimulating activity by the compounds of the invention allows reduced amounts of the growth regulating, cytotoxic, or chemotherapeutic agents thereby potentially lowering the toxicity to the patient.

Cancer is characterized by the increase in the number of abnormal, or neoplastic, cells derived from a normal tissue which proliferate to form a tumor mass, the invasion of adjacent tissues by these neoplastic tumor cells, and the generation of malignant cells which eventually spread via the blood or lymphatic system to regional lymph nodes and to distant sites (metastasis). In a cancerous state a cell proliferates under conditions in which normal cells would not grow. Cancer manifests itself in a wide variety of forms, characterized by different degrees of invasiveness and aggressiveness.

Alteration of gene expression is intimately related to the uncontrolled cell growth and de-differentiation which are a common feature of all cancers. The genomes of certain well studied tumors have been found to show decreased expression of recessive genes, usually referred to as tumor suppression genes, which would normally function to prevent malignant cell growth, and/or overexpression of certain dominant genes, such as oncogenes, that act to promote malignant growth. Each of these genetic changes appears to be responsible for importing some of the traits that, in aggregate, represent the full neoplastic phenotype (Hunter, Cell 64, 1129 [1991]; Bishop, Cell 64, 235-248 [1991]).

A well known mechanism of gene (e.g. oncogene) overexpression in cancer cells is gene amplification. This is a process where in the chromosome of the ancestral cell multiple copies of a particular gene are produced. The process involves unscheduled replication of the region of chromosome comprising the gene, followed by recombination of the replicated segments back into the chromosome (Alitalo *et al.*, Adv. Cancer Res. 47, 235-281 [1986]). It is believed that the overexpression of the gene parallels gene amplification, i.e. is proportionate to the number of copies made.

Proto-oncogenes that encode growth factors and growth factor receptors have been identified to play important roles in the pathogenesis of various human malignancies, including breast cancer. For example, it



has been found that the human ErbB2 gene (*erbB2*, also known as *her2*, or *c-erbB-2*), which encodes a 185-kd transmembrane glycoprotein receptor (p185<sup>HER2</sup>; HER2) related to the epidermal growth factor receptor (EGFR), is overexpressed in about 25% to 30% of human breast cancer (Slamon *et al.*, *Science* 235:177-182 [1987]; Slamon *et al.*, *Science* 244:707-712 [1989]).

5 It has been reported that gene amplification of a protooncogene is an event typically involved in the more malignant forms of cancer, and could act as a predictor of clinical outcome (Schwab *et al.*, *Genes Chromosomes Cancer* 1, 181-193 [1990]; Alitalo *et al.*, *supra*). Thus, *erbB2* overexpression is commonly regarded as a predictor of a poor prognosis, especially in patients with primary disease that involves axillary lymph nodes (Slamon *et al.*, [1987] and [1989], *supra*; Ravdin and Chamness, *Gene* 159:19-27 [1995]; and  
10 Hynes and Stern, *Biochim Biophys Acta* 1198:165-184 [1994]), and has been linked to sensitivity and/or resistance to hormone therapy and chemotherapeutic regimens, including CMF (cyclophosphamide, methotrexate, and fluoruracil) and anthracyclines (Baselga *et al.*, *Oncology* 11(3 Suppl 1):43-48 [1997]). However, despite the association of *erbB2* overexpression with poor prognosis, the odds of HER2-positive patients responding clinically to treatment with taxanes were greater than three times those of HER2-negative  
15 patients (*Ibid*). A recombinant humanized anti-ErbB2 (anti-HER2) monoclonal antibody (a humanized version of the murine anti-ErbB2 antibody 4D5, referred to as rhuMAb HER2 or Herceptin<sup>®</sup>) has been clinically active in patients with ErbB2-overexpressing metastatic breast cancers that had received extensive prior anticancer therapy. (Baselga *et al.*, *J. Clin. Oncol.* 14:737-744 [1996]).

The compounds of the invention may be administered as adjuvants in the treatment of cancers in  
20 which one or more genes in cancer cells are amplified. Gene amplification is a quantitative modification, whereby the actual number of complete coding sequence, i.e. a gene, increases, leading to an increased number of available templates for transcription, an increased number of translatable transcripts, and, ultimately, to an increased abundance of the protein encoded by the amplified gene.

The phenomenon of gene amplification and its underlying mechanisms have been investigated *in*  
25 *vitro* in several prokaryotic and eukaryotic culture systems. The best-characterized example of gene amplification involves the culture of eukaryotic cells in medium containing variable concentrations of the cytotoxic drug methotrexate (MTX). MTX is a folic acid analogue and interferes with DNA synthesis by blocking the enzyme dihydrofolate reductase (DHFR). During the initial exposure to low concentrations of MTX most cells (>99.9%) will die. A small number of cells survive, and are capable of growing in increasing  
30 concentrations of MTX by producing large amounts of DHFR-RNA and protein. The basis of this overproduction is the amplification of the single DHFR gene. The additional copies of the gene are found as extrachromosomal copies in the form of small, supernumerary chromosomes (double minutes) or as integrated chromosomal copies.

Gene amplification is most commonly encountered in the development of resistance to cytotoxic  
35 drugs (antibiotics for bacteria and chemotherapeutic agents for eukaryotic cells) and neoplastic transformation. Transformation of a eukaryotic cell as a spontaneous event or due to a viral or chemical/environmental insult is typically associated with changes in the genetic material of that cell. One of the most common genetic changes observed in human malignancies are mutations of the p53 protein. p53 controls the transition of cells from the stationary (G1) to the replicative (S) phase and prevents this transition  
40 in the presence of DNA damage. In other words, one of the main consequences of disabling p53 mutations is the accumulation and propagation of DNA damage, i.e. genetic changes. Common types of genetic changes

in neoplastic cells are, in addition to point mutations, amplifications and gross, structural alterations, such as translocations.

The amplification of DNA sequences may indicate specific functional requirement as illustrated in the DHFR experimental system. Therefore, the amplification of certain oncogenes in malignancies points toward a causative role of these genes in the process of malignant transformation and maintenance of the transformed phenotype. This hypothesis has gained support in recent studies. For example, the *bcl-2* protein was found to be amplified in certain types of non-Hodgkin's lymphoma. This protein inhibits apoptosis and leads to the progressive accumulation of neoplastic cells. Members of the gene family of growth factor receptors have been found to be amplified in various types of cancers suggesting that overexpression of these receptors may make neoplastic cells less susceptible to limiting amounts of available growth factor. Examples include the amplification of the androgen receptor in recurrent prostate cancer during androgen deprivation therapy and the amplification of the growth factor receptor homologue ERB2 in breast cancer. Lastly, genes involved in intracellular signaling and control of cell cycle progression can undergo amplification during malignant transformation. This is illustrated by the amplification of the *bcl-1* and *ras* genes in various epithelial and lymphoid neoplasms.

These earlier studies illustrate the feasibility of identifying amplified DNA sequences in neoplasms, because this approach can identify genes important for malignant transformation. The case of ERB2 also demonstrates the feasibility from a therapeutic standpoint, since transforming proteins may represent novel and specific targets for tumor therapy.

Several different techniques can be used to demonstrate amplified genomic sequences. Classical cytogenetic analysis of chromosome spreads prepared from cancer cells is adequate to identify gross structural alterations, such as translocations, deletions and inversions. Amplified genomic regions can only be visualized, if they involve large regions with high copy numbers or are present as extrachromosomal material. While cytogenetics was the first technique to demonstrate the consistent association of specific chromosomal changes with particular neoplasms, it is inadequate for the identification and isolation of manageable DNA sequences. The more recently developed technique of comparative genomic hybridization (CGH) has illustrated the widespread phenomenon of genomic amplification in neoplasms. Tumor and normal DNA are hybridized simultaneously onto metaphases of normal cells and the entire genome can be screened by image analysis for DNA sequences that are present in the tumor at an increased frequency. (WO 93/18,186; Gray *et al.*, Radiation Res. 137, 275-289 [1994]). As a screening method, this type of analysis has revealed a large number of recurring amplicons (a stretch of amplified DNA) in a variety of human neoplasms. Although CGH is more sensitive than classical cytogenetic analysis in identifying amplified stretches of DNA, it does not allow a rapid identification and isolation of coding sequences within the amplicon by standard molecular genetic techniques.

The most sensitive methods to detect gene amplification are polymerase chain reaction (PCR)-based assays. These assays utilize very small amount of tumor DNA as starting material, are exquisitely sensitive, provide DNA that is amenable to further analysis, such as sequencing and are suitable for high-volume throughput analysis.

The above-mentioned assays are not mutually exclusive, but are frequently used in combination to identify amplifications in neoplasms. While cytogenetic analysis and CGH represent screening methods to survey the entire genome for amplified regions, PCR-based assays are most suitable for the final identification of coding sequences, i.e. genes in amplified regions. Such genes can be identified by quantitative PCR (S.

Gelmini *et al.*, *Clin. Chem.* 43, 752 [1997]), by comparing DNA from a variety of primary tumors, including breast, lung, colon, prostate, brain, liver, kidney, pancreas, spleen, thymus, testis, ovary, uterus, etc. tumor, or tumor cell lines, with pooled DNA from healthy donors. Quantitative PCR may be performed using a TaqMan instrument (ABI). Gene-specific primers and fluorogenic probes are designed based upon the coding  
 5 sequences of the DNAs.

The compounds of the invention can be used as immunoadjuvants in the treatment of cancers in which amplified genes have been found in cancer cell lines, such as:

Human lung carcinoma cell lines including A549 (SRCC768), Calu-1 (SRCC769), Calu-6 (SRCC770), H157 (SRCC771), H441 (SRCC772), H460 (SRCC773), SKMES-1 (SRCC774) and SW900  
 10 (SRCC775), all available from ATCC. Primary human lung tumor cells usually derive from adenocarcinomas, squamous cell carcinomas, large cell carcinomas, non-small cell carcinomas, small cell carcinomas, and broncho alveolar carcinomas, and include, for example, SRCC724 (squamous cell carcinoma abbreviated as "SqCCa"), SRCC725 (non-small cell carcinoma, abbreviated as "NSCCa"), SRCC726 (adenocarcinoma, abbreviated as "AdenoCa"), SRCC727 (adenocarcinoma), SRCC728 (squamous  
 15 cell carcinoma), SRCC729 (adenocarcinoma), SRCC730 (adeno/squamous cell carcinoma), SRCC731 (adenocarcinoma), SRCC732 (squamous cell carcinoma), SRCC733 (adenocarcinoma), SRCC734 (adenocarcinoma), SRCC735 (broncho alveolar carcinoma, abbreviated as "BAC"), SRCC736 (squamous cell carcinoma), SRCC738 (squamous cell carcinoma), SRCC739 (squamous cell carcinoma), SRCC740 (squamous cell carcinoma), SRCC740 (lung cell carcinoma, abbreviated as "LCCa");

20 Colon cancer cell lines including, for example, ATCC cell lines SW480 (adenocarcinoma, SRCC776), SW620 (lymph node metastasis of colon adenocarcinoma, SRCC777), COLO320 (adenocarcinoma, SRCC778), HT29 (adenocarcinoma, SRCC779), HM7 (carcinoma, SRCC780), CaWiDr (adenocarcinoma, srcc781), HCT116 (carcinoma, SRCC782), SKCO1 (adenocarcinoma, SRCC783), SW403 (adenocarcinoma, SRCC784), LS174T (carcinoma, SRCC785), and HM7 (a high mucin producing variant of  
 25 ATCC colon adenocarcinoma cell line LS 174T, obtained from Dr. Robert Warren, UCSF). Primary colon tumors include colon adenocarcinomas designated CT2 (SRCC742), CT3 (SRCC743), CT8 (SRCC744), CT10 (SRCC745), CT12 (SRCC746), CT14 (SRCC747), CT15 (SRCC748), CT17 (SRCC750), CT1 (SRCC751), CT4 (SRCC752), CT5 (SRCC753), CT6 (SRCC754), CT7 (SRCC755), CT9 (SRCC756), CT11 (SRCC757), CT18 (SRCC758), and DcR3, BACrev, BACfwd, T160, and T159; and

30 Human breast carcinoma cell lines including, for example, HBL100 (SRCC759), MB435s (SRCC760), T47D (SRCC761), MB468 (SRCC762), MB175 (SRCC763), MB361 (SRCC764), BT20 (SRCC765), MCF7 (SRCC766), SKBR3 (SRCC767).

#### 6. Screening Assays for Drug Candidates

Screening assays for drug candidates are designed to identify compounds that bind or complex with  
 35 the polypeptides encoded by the genes identified herein or a biologically active fragment thereof, or otherwise interfere with the interaction of the encoded polypeptides with other cellular proteins. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds, including peptides, preferably soluble peptides, (poly)peptide-  
 40 immunoglobulin fusions, and, in particular, antibodies including, without limitation, poly- and monoclonal antibodies and antibody fragments, single-chain antibodies, anti-idiotypic antibodies, and chimeric or humanized versions of such antibodies or fragments, as well as human antibodies and antibody fragments.

The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

All assays are common in that they call for contacting the drug candidate with a polypeptide encoded by a nucleic acid identified herein under conditions and for a time sufficient to allow these two components to  
5 interact.

In binding assays, the interaction is binding and the complex formed can be isolated or detected in the reaction mixture. In a particular embodiment, the polypeptide encoded by the gene identified herein or the drug candidate is immobilized on a solid phase, e.g. on a microtiter plate, by covalent or non-covalent attachments. Non-covalent attachment generally is accomplished by coating the solid surface with a solution  
10 of the polypeptide and drying. Alternatively, an immobilized antibody, e.g. a monoclonal antibody, specific for the polypeptide to be immobilized can be used to anchor it to a solid surface. The assay is performed by adding the non-immobilized component, which may be labeled by a detectable label, to the immobilized component, e.g. the coated surface containing the anchored component. When the reaction is complete, the non-reacted components are removed, e.g. by washing, and complexes anchored on the solid surface are  
15 detected. When the originally non-immobilized component carries a detectable label, the detection of label immobilized on the surface indicates that complexing occurred. Where the originally non-immobilized component does not carry a label, complexing can be detected, for example, by using a labelled antibody specifically binding the immobilized complex.

If the candidate compound interacts with but does not bind to a particular protein encoded by a gene  
20 identified herein, its interaction with that protein can be assayed by methods well known for detecting protein-protein interactions. Such assays include traditional approaches, such as, cross-linking, co-immunoprecipitation, and co-purification through gradients or chromatographic columns. In addition, protein-protein interactions can be monitored by using a yeast-based genetic system described by Fields and co-workers [Fields and Song, Nature (London) 340, 245-246 (1989); Chien *et al.*, Proc. Natl. Acad. Sci. USA  
25 88, 9578-9582 (1991)] as disclosed by Chevray and Nathans [Proc. Natl. Acad. Sci. USA 89, 5789-5793 (1991)]. Many transcriptional activators, such as yeast GAL4, consist of two physically discrete modular domains, one acting as the DNA-binding domain, while the other one functioning as the transcription activation domain. The yeast expression system described in the foregoing publications (generally referred to as the "two-hybrid system") takes advantage of this property, and employs two hybrid proteins, one in which  
30 the target protein is fused to the DNA-binding domain of GAL4, and another, in which candidate activating proteins are fused to the activation domain. The expression of a GAL1-*lacZ* reporter gene under control of a GAL4-activated promoter depends on reconstitution of GAL4 activity via protein-protein interaction. Colonies containing interacting polypeptides are detected with a chromogenic substrate for  $\beta$ -galactosidase. A complete kit (MATCHMAKER<sup>TM</sup>) for identifying protein-protein interactions between two specific  
35 proteins using the two-hybrid technique is commercially available from Clontech. This system can also be extended to map protein domains involved in specific protein interactions as well as to pinpoint amino acid residues that are crucial for these interactions.

In order to find compounds that interfere with the interaction of a gene identified herein and other intra- or extracellular components can be tested, a reaction mixture is usually prepared containing the product  
40 of the gene and the intra- or extracellular component under conditions and for a time allowing for the interaction and binding of the two products. To test the ability of a test compound to inhibit binding, the

reaction is run in the absence and in the presence of the test compound. In addition, a placebo may be added to a third reaction mixture, to serve as positive control. The binding (complex formation) between the test compound and the intra- or extracellular component present in the mixture is monitored as described above. The formation of a complex in the control reaction(s) but not in the reaction mixture containing the test compound indicates that the test compound interferes with the interaction of the test compound and its reaction partner.

#### 7. Compositions and Methods for the Treatment of Immune Related Diseases

The compositions useful in the treatment of immune related diseases include, without limitation, antibodies, small organic and inorganic molecules, peptides, phosphopeptides, antisense and ribozyme molecules, triple helix molecules, etc. that inhibit or stimulate immune function, for example, T cell proliferation/activation, lymphokine release, or immune cell infiltration.

For example, antisense RNA and RNA molecule act to directly block the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. When antisense DNA is used, oligodeoxyribonucleotides derived from the translation initiation site, e.g. between about -10 and +10 positions of the target gene nucleotide sequence, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. Ribozymes act by sequence-specific hybridization to the complementary target RNA, followed by endonucleolytic cleavage. Specific ribozyme cleavage sites within a potential RNA target can be identified by known techniques. For further details see, e.g. Rossi, Current Biology 4, 469-471 (1994), and PCT publication No. WO 97/33551 (published September 18, 1997).

Nucleic acid molecules in triple helix formation used to inhibit transcription should be single-stranded and composed of deoxynucleotides. The base composition of these oligonucleotides is designed such that it promotes triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of purines or pyrimidines on one strand of a duplex. For further details see, e.g. PCT publication No. WO 97/33551, *supra*.

These molecules can be identified by any or any combination of the screening assays discussed above and/or by any other screening techniques well known for those skilled in the art.

#### 8. Antibodies

Some of the most promising drug candidates according to the present invention are antibodies and antibody fragments which may inhibit (antagonists) or stimulate (agonists) T cell proliferation, eosinophil infiltration, etc.

##### i. Polyclonal Antibodies

Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the polypeptide of the invention or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

ii. Monoclonal Antibodies

Antibodies which recognize and bind to the polypeptides of the invention or which act as agonist therefor may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the polypeptide of the invention or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the polypeptide of the invention or having similar activity as the polypeptide of the invention. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxyapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., *supra*] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies are preferably monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

*In vitro* methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

### iii. Human and Humanized Antibodies

The antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-329 (1988); and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These

non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and coworkers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeven et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the  
5 corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

10 Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985); Boerner et al., J. Immunol., 147(1):86-95 (1991); U. S. 5,750, 373]. Similarly, human antibodies can be  
15 made by introducing of human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific  
20 publications: Marks et al., Bio/Technology 10, 779-783 (1992); Lonberg et al., Nature 368 856-859 (1994); Morrison, Nature 368, 812-13 (1994); Fishwild et al., Nature Biotechnology 14, 845-51 (1996); Neuberger, Nature Biotechnology 14, 826 (1996); Lonberg and Huszar, Intern. Rev. Immunol. 13 65-93 (1995).

#### iv Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding  
25 specificities for at least two different antigens. In the present case, one of the binding specificities may be for the polypeptide of the invention, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the coexpression of two immunoglobulin heavy-chain/light-  
30 chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 [1983]). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in  
35 Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain  
40 binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are cotransfected



into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

v Heteroconjugate Antibodies

Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

vi Effector function engineering

It may be desirable to modify the antibody of the invention with respect to effector function, so as to enhance the effectiveness of the antibody in treating an immune related disease, for example. For example cysteine residue(s) may be introduced in the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated may have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp. Med. 176:1191-1195 (1992) and Shopes, B. J. Immunol. 148:2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity may also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research 53:2560-2565 (1993). Alternatively, an antibody can be engineered which has dual Fc regions and may thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design 3:219-230 (1989).

vii Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g. an enzymatically active toxin of bacterial, fungal, plant or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof which can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, *Aleurites fordii* proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include <sup>212</sup>Bi, <sup>131</sup>I, <sup>131</sup>In, <sup>90</sup>Y and <sup>186</sup>Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene

triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody may be conjugated to a "receptor" (such streptavidin) for utilization in tissue pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g. avidin) which is conjugated to a cytotoxic agent (e.g. a radionucleotide).

#### viii Immunoliposomes

The proteins, antibodies, etc. disclosed herein may also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein *et al.*, Proc. Natl. Acad. Sci. USA, 82:3688 (1985); Hwang *et al.*, Proc. Natl. Acad. Sci. USA, 77:4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin *et al.*, J. Biol. Chem. 257: 286-288 (1982) via a disulfide interchange reaction. A chemotherapeutic agent (such as doxorubicin) may be optionally contained within the liposome. See Gabizon *et al.*, J. National Cancer Inst. 81(19)1484 (1989).

#### 9. Pharmaceutical Compositions

The active molecules of the invention, polypeptides and antibodies, as well as other molecules identified by the screening assays disclosed above, can be administered for the treatment of immune related diseases, in the form of pharmaceutical compositions.

Therapeutic formulations of the active molecule, preferably a polypeptide or antibody of the invention, are prepared for storage by mixing the active molecule having the desired degree of purity with optional pharmaceutically acceptable carriers, excipients or stabilizers (*Remington's Pharmaceutical Sciences* 16th edition, Osol, A. Ed. [1980]), in the form of lyophilized formulations or aqueous solutions. Acceptable carriers, excipients, or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride, benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (e.g. Zn-protein complexes); and/or non-ionic surfactants such as TWEEN<sup>TM</sup>, PLURONICS<sup>TM</sup> or polyethylene glycol (PEG).

Compounds identified by the screening assays of the present invention can be formulated in an analogous manner, using standard techniques well known in the art.

Lipofections or liposomes can also be used to deliver the polypeptide, antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment which specifically

binds to the binding domain of the target protein is preferred. For example, based upon the variable region sequences of an antibody, peptide molecules can be designed which retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology (see, e.g. Marasco *et al.*, Proc. Natl. Acad. Sci. USA 90, 7889-7893 [1993]).

5       The formulation herein may also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition may comprise a cytotoxic agent, cytokine or growth inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

10       The active molecules may also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles and nanocapsules) or in macroemulsions. Such techniques are disclosed in *Remington's Pharmaceutical Sciences* 16th edition, Osol, A. Ed. (1980).

15       The formulations to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations may be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g. films, or microcapsules. Examples of sustained-release matrices include  
20   polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and  $\gamma$  ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT<sup>TM</sup> (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable  
25   release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods. When encapsulated antibodies remain in the body for a long time, they may denature or aggregate as a result of exposure to moisture at 37C, resulting in a loss of biological activity and possible changes in immunogenicity.

Rational strategies can be devised for stabilization depending on the mechanism involved. For example, if the aggregation mechanism is discovered to be intermolecular S-S bond formation through thio-disulfide  
30   interchange, stabilization may be achieved by modifying sulfhydryl residues, lyophilizing from acidic solutions, controlling moisture content, using appropriate additives, and developing specific polymer matrix compositions.

#### 10.    Methods of Treatment

It is contemplated that the polypeptides, antibodies and other active compounds of the present  
35   invention may be used to treat various immune related diseases and conditions, such as T cell mediated diseases, including those characterized by infiltration of inflammatory cells into a tissue, stimulation of T-cell proliferation, inhibition of T-cell proliferation, increased or decreased vascular permeability or the inhibition thereof.

Exemplary conditions or disorders to be treated with the polypeptides, antibodies and other  
40   compounds of the invention, include, but are not limited to systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic arthritis, spondyloarthropathies, systemic sclerosis (scleroderma), idiopathic inflammatory

myopathies (dermatomyositis, polymyositis), Sjögren's syndrome, systemic vasculitis, sarcoidosis, autoimmune hemolytic anemia (immune pancytopenia, paroxysmal nocturnal hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic purpura, immune-mediated thrombocytopenia), thyroiditis (Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic thyroiditis), diabetes mellitus, immune-mediated renal disease (glomerulonephritis, tubulointerstitial nephritis), demyelinating diseases of the central and peripheral nervous systems such as multiple sclerosis, idiopathic demyelinating polyneuropathy or Guillain-Barré syndrome, and chronic inflammatory demyelinating polyneuropathy, hepatobiliary diseases such as infectious hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune chronic active hepatitis, primary biliary cirrhosis, granulomatous hepatitis, and sclerosing cholangitis, inflammatory and fibrotic lung diseases such as inflammatory bowel disease (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and Whipple's disease, autoimmune or immune-mediated skin diseases including bullous skin diseases, erythema multiforme and contact dermatitis, psoriasis, allergic diseases such as asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity and urticaria, immunologic diseases of the lung such as eosinophilic pneumonias, idiopathic pulmonary fibrosis and hypersensitivity pneumonitis, transplantation associated diseases including graft rejection and graft - versus-host-disease.

In systemic lupus erythematosus, the central mediator of disease is the production of auto-reactive antibodies to self proteins/tissues and the subsequent generation of immune-mediated inflammation. antibodies either directly or indirectly mediate tissue injury. Though T lymphocytes have not been shown to be directly involved in tissue damage, T lymphocytes are required for the development of auto-reactive antibodies. The genesis of the disease is thus T lymphocyte dependent. Multiple organs and systems are affected clinically including kidney, lung, musculoskeletal system, mucocutaneous, eye, central nervous system, cardiovascular system, gastrointestinal tract, bone marrow and blood.

Rheumatoid arthritis (RA) is a chronic systemic autoimmune inflammatory disease that mainly involves the synovial membrane of multiple joints with resultant injury to the articular cartilage. The pathogenesis is T lymphocyte dependent and is associated with the production of rheumatoid factors, auto-antibodies directed against self IgG, with the resultant formation of immune complexes that attain high levels in joint fluid and blood. These complexes in the joint may induce the marked infiltrate of lymphocytes and monocytes into the synovium and subsequent marked synovial changes; the joint space/fluid if infiltrated by similar cells with the addition of numerous neutrophils. Tissues affected are primarily the joints, often in symmetrical pattern. However, extra-articular disease also occurs in two major forms. One form is the development of extra-articular lesions with ongoing progressive joint disease and typical lesions of pulmonary fibrosis, vasculitis, and cutaneous ulcers. The second form of extra-articular disease is the so called Felty's syndrome which occurs late in the RA disease course, sometimes after joint disease has become quiescent, and involves the presence of neutropenia, thrombocytopenia and splenomegaly. This can be accompanied by vasculitis in multiple organs with formations of infarcts, skin ulcers and gangrene. Patients often also develop rheumatoid nodules in the subcutis tissue overlying affected joints; the nodules late stage have necrotic centers surrounded by a mixed inflammatory cell infiltrate. Other manifestations which can occur in RA include: pericarditis, pleuritis, coronary arteritis, interstitial pneumonitis with pulmonary fibrosis, keratoconjunctivitis sicca, and rheumatoid nodules.

Juvenile chronic arthritis is a chronic idiopathic inflammatory disease which begins often at less than 16 years of age. Its phenotype has some similarities to RA; some patients which are rheumatoid factor positive

are classified as juvenile rheumatoid arthritis. The disease is sub-classified into three major categories: pauciarticular, polyarticular, and systemic. The arthritis can be severe and is typically destructive and leads to joint ankylosis and retarded growth. Other manifestations can include chronic anterior uveitis and systemic amyloidosis.

- 5 Spondyloarthropathies are a group of disorders with some common clinical features and the common association with the expression of HLA-B27 gene product. The disorders include: ankylosing spondylitis, Reiter's syndrome (reactive arthritis), arthritis associated with inflammatory bowel disease, spondylitis associated with psoriasis, juvenile onset spondyloarthropathy and undifferentiated spondyloarthropathy. Distinguishing features include sacroileitis with or without spondylitis; inflammatory asymmetric arthritis; 10 association with HLA-B27 (a serologically defined allele of the HLA-B locus of class I MHC); ocular inflammation, and absence of autoantibodies associated with other rheumatoid disease. The cell most implicated as key to induction of the disease is the CD8+ T lymphocyte, a cell which targets antigen presented by class I MHC molecules. CD8+ T cells may react against the class I MHC allele HLA-B27 as if it were a foreign peptide expressed by MHC class I molecules. It has been hypothesized that an epitope of HLA-B27 15 may mimic a bacterial or other microbial antigenic epitope and thus induce a CD8+ T cells response.

- Systemic sclerosis (scleroderma) has an unknown etiology. A hallmark of the disease is induration of the skin; likely this is induced by an active inflammatory process. Scleroderma can be localized or systemic; vascular lesions are common and endothelial cell injury in the microvasculature is an early and important event in the development of systemic sclerosis; the vascular injury may be immune mediated. An 20 immunologic basis is implied by the presence of mononuclear cell infiltrates in the cutaneous lesions and the presence of anti-nuclear antibodies in many patients. ICAM-1 is often upregulated on the cell surface of fibroblasts in skin lesions suggesting that T cell interaction with these cells may have a role in the pathogenesis of the disease. Other organs involved include: the gastrointestinal tract: smooth muscle atrophy and fibrosis resulting in abnormal peristalsis/motility; kidney: concentric subendothelial intimal proliferation 25 affecting small arcuate and interlobular arteries with resultant reduced renal cortical blood flow, results in proteinuria, azotemia and hypertension; skeletal muscle: atrophy, interstitial fibrosis; inflammation; lung: interstitial pneumonitis and interstitial fibrosis; and heart: contraction band necrosis, scarring/fibrosis.

- Idiopathic inflammatory myopathies including dermatomyositis, polymyositis and others are disorders of chronic muscle inflammation of unknown etiology resulting in muscle weakness. Muscle 30 injury/inflammation is often symmetric and progressive. Autoantibodies are associated with most forms. These myositis-specific autoantibodies are directed against and inhibit the function of components, proteins and RNA's, involved in protein synthesis.

- Sjogren's syndrome is due to immune-mediated inflammation and subsequent functional destruction of the tear glands and salivary glands. The disease can be associated with or accompanied by inflammatory 35 connective tissue diseases. The disease is associated with autoantibody production against Ro and La antigens, both of which are small RNA-protein complexes. Lesions result in keratoconjunctivitis sicca, xerostomia, with other manifestations or associations including biliary cirrhosis, peripheral or sensory neuropathy, and palpable purpura.

- Systemic vasculitis are diseases in which the primary lesion is inflammation and subsequent damage 40 to blood vessels which results in ischemia/necrosis/degeneration to tissues supplied by the affected vessels and eventual end-organ dysfunction in some cases. Vasculitides can also occur as a secondary lesion or sequelae to other immune-inflammatory mediated diseases such as rheumatoid arthritis, systemic sclerosis,

etc., particularly in diseases also associated with the formation of immune complexes. Diseases in the primary systemic vasculitis group include: systemic necrotizing vasculitis: polyarteritis nodosa, allergic angiitis and granulomatosis, polyangiitis; Wegener's granulomatosis; lymphomatoid granulomatosis; and giant cell arteritis. Miscellaneous vasculitides include: mucocutaneous lymph node syndrome (MLNS or  
5 Kawasaki's disease), isolated CNS vasculitis, Behet's disease, thromboangiitis obliterans (Buerger's disease) and cutaneous necrotizing venulitis. The pathogenic mechanism of most of the types of vasculitis listed is believed to be primarily due to the deposition of immunoglobulin complexes in the vessel wall and subsequent induction of an inflammatory response either via ADCC, complement activation, or both.

Sarcoidosis is a condition of unknown etiology which is characterized by the presence of epithelioid  
10 granulomas in nearly any tissue in the body; involvement of the lung is most common. The pathogenesis involves the persistence of activated macrophages and lymphoid cells at sites of the disease with subsequent chronic sequelae resultant from the release of locally and systemically active products released by these cell types.

Autoimmune hemolytic anemia including autoimmune hemolytic anemia, immune pancytopenia, and  
15 paroxysmal nocturnal hemoglobinuria is a result of production of antibodies that react with antigens expressed on the surface of red blood cells (and in some cases other blood cells including platelets as well) and is a reflection of the removal of those antibody coated cells via complement mediated lysis and/or ADCC/Fc-receptor-mediated mechanisms.

In autoimmune thrombocytopenia including thrombocytopenic purpura, and immune-mediated  
20 thrombocytopenia in other clinical settings, platelet destruction/removal occurs as a result of either antibody or complement attaching to platelets and subsequent removal by complement lysis, ADCC or FC-receptor mediated mechanisms.

Thyroiditis including Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, and atrophic thyroiditis, are the result of an autoimmune response against thyroid antigens with production of  
25 antibodies that react with proteins present in and often specific for the thyroid gland. Experimental models exist including spontaneous models: rats (BUF and BB rats) and chickens (obese chicken strain); inducible models: immunization of animals with either thyroglobulin, thyroid microsomal antigen (thyroid peroxidase).

Type I diabetes mellitus or insulin-dependent diabetes is the autoimmune destruction of pancreatic islet  $\beta$  cells; this destruction is mediated by auto-antibodies and auto-reactive T cells. Antibodies to insulin or  
30 the insulin receptor can also produce the phenotype of insulin-non-responsiveness.

Immune mediated renal diseases, including glomerulonephritis and tubulointerstitial nephritis, are the result of antibody or T lymphocyte mediated injury to renal tissue either directly as a result of the production of autoreactive antibodies or T cells against renal antigens or indirectly as a result of the deposition of antibodies and/or immune complexes in the kidney that are reactive against other, non-renal  
35 antigens. Thus other immune-mediated diseases that result in the formation of immune-complexes can also induce immune mediated renal disease as an indirect sequelae. Both direct and indirect immune mechanisms result in inflammatory response that produces/induces lesion development in renal tissues with resultant organ function impairment and in some cases progression to renal failure. Both humoral and cellular immune mechanisms can be involved in the pathogenesis of lesions.

Demyelinating diseases of the central and peripheral nervous systems, including Multiple Sclerosis;  
40 idiopathic demyelinating polyneuropathy or Guillain-Barr syndrome; and Chronic Inflammatory Demyelinating Polyneuropathy, are believed to have an autoimmune basis and result in nerve demyelination

as a result of damage caused to oligodendrocytes or to myelin directly. In MS there is evidence to suggest that disease induction and progression is dependent on T lymphocytes. Multiple Sclerosis is a demyelinating disease that is T lymphocyte-dependent and has either a relapsing-remitting course or a chronic progressive course. The etiology is unknown; however, viral infections, genetic predisposition, environment, and autoimmunity all contribute. Lesions contain infiltrates of predominantly T lymphocyte mediated, microglial cells and infiltrating macrophages; CD4+T lymphocytes are the predominant cell type at lesions. The mechanism of oligodendrocyte cell death and subsequent demyelination is not known but is likely T lymphocyte driven.

Inflammatory and Fibrotic Lung Disease, including Eosinophilic Pneumonias; Idiopathic Pulmonary Fibrosis, and Hypersensitivity Pneumonitis may involve a dysregulated immune-inflammatory response. Inhibition of that response would be of therapeutic benefit.

Autoimmune or Immune-mediated Skin Disease including Bullous Skin Diseases, Erythema Multiforme, and Contact Dermatitis are mediated by auto-antibodies, the genesis of which is T lymphocyte-dependent.

Psoriasis is a T lymphocyte-mediated inflammatory disease. Lesions contain infiltrates of T lymphocytes, macrophages and antigen processing cells, and some neutrophils.

Allergic diseases, including asthma; allergic rhinitis; atopic dermatitis; food hypersensitivity; and urticaria are T lymphocyte dependent. These diseases are predominantly mediated by T lymphocyte induced inflammation, IgE mediated-inflammation or a combination of both.

Transplantation associated diseases, including Graft rejection and Graft-Versus-Host-Disease (GVHD) are T lymphocyte-dependent; inhibition of T lymphocyte function is ameliorative.

Other diseases in which intervention of the immune and/or inflammatory response have benefit are Infectious disease including but not limited to viral infection (including but not limited to AIDS, hepatitis A, B, C, D, E) bacterial infection, fungal infections, and protozoal and parasitic infections (molecules (or derivatives/agonists) which stimulate the MLR can be utilized therapeutically to enhance the immune response to infectious agents), diseases of immunodeficiency (molecules/derivatives/agonists) which stimulate the MLR can be utilized therapeutically to enhance the immune response for conditions of inherited, acquired, infectious induced (as in HIV infection), or iatrogenic (i.e. as from chemotherapy) immunodeficiency), and neoplasia.

It has been demonstrated that some human cancer patients develop an antibody and/or T lymphocyte response to antigens on neoplastic cells. It has also been shown in animal models of neoplasia that enhancement of the immune response can result in rejection or regression of that particular neoplasm. Molecules that enhance the T lymphocyte response in the MLR have utility in vivo in enhancing the immune response against neoplasia. Molecules which enhance the T lymphocyte proliferative response in the MLR (or small molecule agonists or antibodies that affected the same receptor in an agonistic fashion) can be used therapeutically to treat cancer. Molecules that inhibit the lymphocyte response in the MLR also function in vivo during neoplasia to suppress the immune response to a neoplasm; such molecules can either be expressed by the neoplastic cells themselves or their expression can be induced by the neoplasm in other cells. Antagonism of such inhibitory molecules (either with antibody, small molecule antagonists or other means) enhances immune-mediated tumor rejection.

Additionally, inhibition of molecules with proinflammatory properties may have therapeutic benefit in reperfusion injury; stroke; myocardial infarction; atherosclerosis; acute lung injury; hemorrhagic shock; burn; sepsis/septic shock; acute tubular necrosis; endometriosis; degenerative joint disease and pancreatitis.

The compounds of the present invention, e.g. polypeptides or antibodies, are administered to a mammal, preferably a human, in accord with known methods, such as intravenous administration as a bolus or by continuous infusion over a period of time, by intramuscular, intraperitoneal, intracerebrospinal, subcutaneous, intra-articular, intrasynovial, intrathecal, oral, topical, or inhalation (intranasal, intrapulmonary) routes. Intravenous or inhaled administration of polypeptides and antibodies is preferred.

In immunoadjuvant therapy, other therapeutic regimens, such administration of an anti-cancer agent, may be combined with the administration of the proteins, antibodies or compounds of the instant invention. For example, the patient to be treated with a the immunoadjuvant of the invention may also receive an anti-cancer agent (chemotherapeutic agent) or radiation therapy. Preparation and dosing schedules for such chemotherapeutic agents may be used according to manufacturers' instructions or as determined empirically by the skilled practitioner. Preparation and dosing schedules for such chemotherapy are also described in *Chemotherapy Service* Ed., M.C. Perry, Williams & Wilkins, Baltimore, MD (1992). The chemotherapeutic agent may precede, or follow administration of the immunoadjuvant or may be given simultaneously therewith. Additionally, an anti-oestrogen compound such as tamoxifen or an anti-progesterone such as onapristone (see, EP 616812) may be given in dosages known for such molecules.

It may be desirable to also administer antibodies against other immune disease associated or tumor associated antigens, such as antibodies which bind to CD20, CD11a, CD18, ErbB2, EGFR, ErbB3, ErbB4, or vascular endothelial factor (VEGF). Alternatively, or in addition, two or more antibodies binding the same or two or more different antigens disclosed herein may be coadministered to the patient. Sometimes, it may be beneficial to also administer one or more cytokines to the patient. In one embodiment, the polypeptides of the invention are coadministered with a growth inhibitory agent. For example, the growth inhibitory agent may be administered first, followed by a polypeptide of the invention. However, simultaneous administration or administration first is also contemplated. Suitable dosages for the growth inhibitory agent are those presently used and may be lowered due to the combined action (synergy) of the growth inhibitory agent and the polypeptide of the invention.

For the treatment or reduction in the severity of immune related disease, the appropriate dosage of an a compound of the invention will depend on the type of disease to be treated, as defined above, the severity and course of the disease, whether the agent is administered for preventive or therapeutic purposes, previous therapy, the patient's clinical history and response to the compound, and the discretion of the attending physician. The compound is suitably administered to the patient at one time or over a series of treatments.

For example, depending on the type and severity of the disease, about 1 ug/kg to 15 mg/kg (e.g. 0.1-20mg/kg) of polypeptide or antibody is an initial candidate dosage for administration to the patient, whether, for example, by one or more separate administrations, or by continuous infusion. A typical daily dosage might range from about 1 ug/kg to 100 mg/kg or more, depending on the factors mentioned above. For repeated administrations over several days or longer, depending on the condition, the treatment is sustained until a desired suppression of disease symptoms occurs. However, other dosage regimens may be useful. The progress of this therapy is easily monitored by conventional techniques and assays.



### 11. Articles of Manufacture

In another embodiment of the invention, an article of manufacture containing materials useful for the diagnosis or treatment of the disorders described above is provided. The article of manufacture comprises a container and a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers may be formed from a variety of materials such as glass or plastic. The container holds a composition which is effective for diagnosing or treating the condition and may have a sterile access port (for example the container may be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle). The active agent in the composition is usually a polypeptide or an antibody of the invention. The label on, or associated with, the container indicates that the composition is used for diagnosing or treating the condition of choice. The article of manufacture may further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution and dextrose solution. It may further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

### 12. Diagnosis and Prognosis of Immune Related Disease

Cell surface proteins, such as proteins which are overexpressed in certain immune related diseases, are excellent targets for drug candidates or disease treatment. The same proteins along with secreted proteins encoded by the genes amplified in immune related disease states find additional use in the diagnosis and prognosis of these diseases. For example, antibodies directed against the protein products of genes amplified in multiple sclerosis, rheumatoid arthritis, or another immune related disease, can be used as diagnostics or prognostics.

For example, antibodies, including antibody fragments, can be used to qualitatively or quantitatively detect the expression of proteins encoded by amplified or overexpressed genes ("marker gene products"). The antibody preferably is equipped with a detectable, e.g. fluorescent label, and binding can be monitored by light microscopy, flow cytometry, fluorimetry, or other techniques known in the art. These techniques are particularly suitable, if the overexpressed gene encodes a cell surface protein. Such binding assays are performed essentially as described above.

*In situ* detection of antibody binding to the marker gene products can be performed, for example, by immunofluorescence or immunoelectron microscopy. For this purpose, a histological specimen is removed from the patient, and a labeled antibody is applied to it, preferably by overlaying the antibody on a biological sample. This procedure also allows for determining the distribution of the marker gene product in the tissue examined. It will be apparent for those skilled in the art that a wide variety of histological methods are readily available for *in situ* detection.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

### EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Manassas, VA. Unless otherwise noted, the present invention uses standard procedures of recombinant DNA

technology, such as those described hereinabove and in the following textbooks: Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press N.Y., 1989; Ausubel *et al.*, *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y., 1989; Innis *et al.*, *PCR Protocols: A Guide to Methods and Applications*, Academic Press, inc., N.Y., 1990; Harlow *et al.*,  
 5 *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, 1988; Gait, M.J., *Oligonucleotide Synthesis*, IRL Press, Oxford, 1984; R.I. Freshney, *Animal Cell Culture*, 1987; Coligan *et al.*, *Current Protocols in Immunology*, 1991.

## EXAMPLE 1

Isolation of cDNA clones Encoding Human PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or

10 PRO326I. Isolation of cDNA Clones Encoding Human PRO245

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary  
 15 EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul *et al.*, *Methods in Enzymology* 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program phrap (Phil Green, University  
 20 of Washington, Seattle, Washington).

A consensus DNA sequence encoding PRO245 was assembled relative to the other identified EST sequences, wherein the consensus sequence was designated herein as DNA30954 (see Figs. 3A-3B), wherein the polypeptide showed some structural homology to transmembrane protein receptor tyrosine kinase proteins.

25 Based on the DNA30954 consensus sequence, oligonucleotides were synthesized to identify by PCR a cDNA library that contained the sequence of interest and for use as probes to isolate a clone of the full-length coding sequence for PRO245.

A pair of PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-ATCGTTGTGAAGTTAGTGCCCC-3' (SEQ ID NO:4)

30 reverse PCR primer 5'-ACCTGCGATATCCAACAGAATTG-3' (SEQ ID NO:5)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA30954 sequence which had the following nucleotide sequence

hybridization probe

5'-GGAAGAGGATACAGTCACTCTGGAAGTATTAGTGGCTCCAGCAGTTCC-3' (SEQ ID NO:6)

35 In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO245 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal liver tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available  
 40 reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD;

pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO245 [herein designated as UNQ219 (DNA35638)] and the derived protein sequence for PRO245.

5 The entire nucleotide sequence of UNQ219 (DNA35638) is shown in Figure 1 (SEQ ID NO:1). Clone UNQ219 (DNA35638) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 89-91 [Kozak et al., supra] and ending at the stop codon at nucleotide positions 1025-1027 (Fig. 1; SEQ ID NO:1). The predicted polypeptide precursor is 312 amino acids long (Fig. 2). Clone UNQ219 (DNA35638) has been deposited with ATCC on September 17, 1997 and is assigned ATCC  
10 Deposit No. 209265.

Analysis of the amino acid sequence of the full-length PRO245 suggests that a portion of it possesses 60% amino acid identity with the human c-myc protein and, therefore, may be a new member of the transmembrane protein receptor tyrosine kinase family.

## II. Isolation of cDNA clones Encoding PRO217

15 The extracellular domain (ECD) sequences (including the secretion signal sequence, if any) from about 950 known secreted proteins from the Swiss-Prot public database were used to search EST databases. The EST databases included public databases (e.g., Dayhof, GenBank), and proprietary databases (e.g. LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul, SF and Gish (1996), *Methods in Enzymology* 266: 460-80 (1996);  
20 <http://blast.wustl.edu/blast/README.html>) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequences. Those comparisons with a Blast score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, WA; (<http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>)).

25 Consensus DNA sequences encoding EGF-like homologues were assembled (DNA28726, SEQ ID NO: 19, Fig. 7A; DNA28730, SEQ ID NO: 21, Fig. 7B and DNA28760, SEQ ID NO: 20, Fig. 7C) using phrap. In some cases, the consensus DNA sequence was extended using repeated cycles of blast and phrap to extend the consensus sequence as far as possible using the three sources of EST sequences listed above. (Indicated as second alignment figure).

30 Based on this consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence. The pair of forward and reverse PCR primers (notated as \*.f and \*.r, respectively) may range from 20 to 30 nucleotides (typically 24), and are designed to give a PCR product of 100-1000 bp in length. The probe sequences (notated as \*.p) are typically 40-55 bp (typically 50) in length. In some cases additional  
35 oligonucleotides are synthesized when the consensus sequence is greater than 1-1.5 kbp. In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel *et al.*, *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the PCR primers. This library was used to isolate DNA32279, DNA32292 and  
40 DNA33094 was fetal kidney, fetal lung and fetal lung, respectively.

RNA for the construction of the cDNA libraries was isolated using standard isolation protocols, e.g., Ausubel *et al.*, *Current Protocols in Molecular Biology*, from tissue or cell line sources or it was purchased from

commercial sources (e.g., Clontech). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods (e.g., Ausubel *et al.*) using commercially available reagents (e.g., Invitrogen). The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation in a suitable cloning vector (pRK5B or pRK5D) in the unique XhoI and NotI sites.

A cDNA clone was sequenced in its entirety. The entire nucleotide sequence of EGF-like homologues is shown in Figures 5A (SEQ ID NO: 13), 5B (SEQ ID NO: 14) and 5C (SEQ ID NO: 15). The predicted polypeptide is 448, 353, and 379 (PRO217) amino acid in length, respectively, with a molecule weight of approximately 50.15, 38.19 and 41.52 kDa, respectively.

10 The oligonucleotide sequences used in the above procedure were the following:

28726.p (OLI500) (SEQ ID NO: 60)

GGGTACACCTGCTCCTGCACCGACGGATATTGGCTTCTGGAAGGCC

28726.f (OLI 502) (SEQ ID NO: 61)

15 ACAGATTCCCACCAAGTGCAACC

28726.r (OLI 503) (SEQ ID NO: 62)

CACACTCGTTCACATCTTGGC

20 28730.p (OLI 516) (SEQ ID NO: 63)

AGGGAGCACGGACAGTGTGCAGATGTGGACGAGTGCTCACTAGCA

28730.f (OLI 517) (SEQ ID NO: 64)

AGAGTGTATCTCTGGCTACGC

25

28730.r (OLI 518) (SEQ ID NO: 65)

TAAGTCCGGCACATTACAGGTC

28760.p (OLI 617) (SEQ ID NO: 66)

30 CCCACGATGTATGAATGGTGGACTTTGTGTGACTCCTGGTTTCTGCATC

28760.f (OLI 618) (SEQ ID NO: 67)

AAAGACGCATCTGCGAGTGTCC

35 28760.r (OLI 619) (SEQ ID NO: 68)

TGCTGATTTCACTGCTCTCCC

### III. Isolation of cDNA clones Encoding Human PRO301

The extracellular domain (ECD) sequences (including the secretion signal sequence, if any) from about 950 known secreted proteins from the Swiss-Prot public database were used to search EST databases. The EST databases included public EST databases (e.g., GenBank), a proprietary EST database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or

BLAST2 [Altschul et al., Methods in Enzymology, 266:460-480 (1996); <http://blast.wustl.edu/blast/README.html>] as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequences. Those comparisons resulting in a BLAST score of 70 (or in some cases, 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

A consensus DNA sequence encoding DNA35936 was assembled using phrap. In some cases, the consensus DNA sequence was extended using repeated cycles of blast and phrap to extend the consensus sequence as far as possible using the three sources of EST sequences listed above. The extended assembly sequence is indicated as a second alignment figure, as shown in Fig. 17.

Based on this consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence. Forward and reverse PCR primers (notated as \*.f and \*.r, respectively) may range from 20 to 30 nucleotides (typically about 24), and are designed to give a PCR product of 100-1000 bp in length. The probe sequences (notated as \*.p) are typically 40-55 bp (typically about 50) in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than 1-1.5 kbp. In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the in vivo cloning procedure using the probe oligonucleotide and one of the PCR primers.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO301 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal kidney. The cDNA libraries used to isolated the cDNA clones were constructed by standard methods using commercially available reagents (e.g., Invitrogen, San Diego, CA; Clontech, etc.) The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

A cDNA clone was sequenced in its entirety. The full length nucleotide sequence of native sequence PRO301 is shown in Figure 15 (SEQ ID NO: 75). Clone DNA40628 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 52-54 [Kozak et al., *supra*] (Fig. 15; SEQ ID NO: 75). The predicted polypeptide precursor is 299 amino acids long with a predicted molecular weight of 32583 daltons and pI of 8.29. Clone DNA40628 has been deposited with ATCC and is assigned ATCC deposit No. 209432.

Based on a BLAST and FastA sequence alignment analysis of the full-length sequence, PRO301 shows amino acid sequence identity to A33 antigen precursor (30%) and coxsackie and adenovirus receptor protein (29%).

The oligonucleotide sequences used in the above procedure were the following:

OLI2162 (35936.f1) (SEQ ID NO:78)  
TCGCGGAGCTGTGTTCTGTTTCCC

OLI2163 (35936.p1) (SEQ ID NO:79)  
TGATCGCGATGGGGACAAAGGCGCAAGCTCGAGAGGAACTGTTGTGCCT

5 OLI2164 (35936.f2) (SEQ ID NO:80)  
ACACCTGGTTCAAAGATGGG

OLI2165 (35936.r1) (SEQ ID NO:81)  
TAGGAAGAGTTGCTGAAGGCACGG

10

OLI2166 (35936.f3) (SEQ ID NO:82)  
TTGCCTTACTCAGGTGCTAC

OLI2167 (35936.r2) (SEQ ID NO:83)  
15 ACTCAGCAGTGGTAGGAAAG

#### IV. Isolation of cDNA Clones Encoding Human PRO266

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul et al., Methods in Enzymology 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

Based on the expression sequence tag (SEQ ID NO:257) shown in Figure 24, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO266. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the primer pairs.

A pair of PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-GTTGGATCTGGGCAACAATAAC-3' (SEQ ID NO:258)

reverse PCR primer 5'-ATTGTTGTGCAGGCTGAGTTTAAG-3' (SEQ ID NO:259)

40 Additionally, a synthetic oligonucleotide hybridization probe was constructed from SEQ ID NO:257 which had the following nucleotide sequence:

hybridization probe

5'-GGTGGCTATACATGGATAGCAATTACCTGGACACGCTGTCCCGGG-3' (SEQ ID NO: 260)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO266 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal brain tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO266 [herein designated as UNQ233 (DNA37150-seq min)] (SEQ ID NO:236) and the derived protein sequence for PRO266.

The entire nucleotide sequence of UNQ233 (DNA37150-seq min) is shown in Figures 20A and 20B (SEQ ID NO:236). Clone UNQ233 (DNA37150-seq min) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 1-3 [Kozak et al., *supra*] and ending at the stop codon after nucleotide position 2088 of SEQ ID NO: 237. The predicted polypeptide precursor is 696 amino acids long (Figure 21). Clone UNQ233 (DNA37150-seq min) has been deposited with ATCC and is assigned ATCC deposit no. 209401.

Analysis of the amino acid sequence of the full-length PRO266 polypeptide suggests that portions of it possess significant homology to the SLIT protein as shown in Figures 22A-22D and 23A-23D, thereby indicating that PRO266 may be a novel leucine rich repeat protein.

## V. Isolation of cDNA Clones Encoding Human PRO335, PRO331 or PRO326

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul et al., *Methods in Enzymology* 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. This consensus sequence is herein designated SEQ ID NO:264, see Figures 27A and 27B.

Based on the SEQ ID NO:264 consensus sequence, and SEQ ID NO:286, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO335, PRO331 or PRO326. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In

order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the primer pairs.

- 5 A number of PCR primers (forward and reverse) were synthesized as shown in Figure 30C and Figure 31 (forward SEQ ID NOS:271-274; reverse SEQ ID NOS:275-277) and yet another primer, SEQ ID NO:278 shown in Figure 31 for determination of PRO335. For determination of PRO40981, the primers are shown in Figure 36, (forward is SEQ ID NO:295; reverse is SEQ ID NO:296; and the other is SEQ ID NO:297). For the determination of PRO326, a 5' splice variant of PRO335, the primers used are shown in  
10 Figures 40 and Figures 41.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO335, PRO331 or PRO326 gene using the probe oligonucleotide and one of the PCR primers.

- 15 RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue (PRO335 and PRO326) and human fetal brain (PRO331). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined  
20 orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

- DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO335, PRO331 or PRO326 [herein designated as SEQ ID NOS:261, 279 or 298, and the derived protein  
25 sequence for PRO335, PRO331 or PRO326.

The entire nucleotide sequences are shown in Figures 25A-B, 32 and 37A-C. The nucleic acid shown in Figure 32 has been deposited with the ATCC on 7 November 1997 and is assigned ATCC Accession No. 209439.

- Analysis of the amino acid sequence of the full-length PRO335, PRO331 or PRO326 polypeptide  
30 suggests that portions of it possess significant homology to the LIG-1 protein as shown in Figures 28A-28C, 34A-34E and 39A-39D, thereby indicating that PRO335, PRO331 and PRO326 may be a novel LIG-1-related protein.

## EXAMPLE 2

### Stimulatory Activity in Mixed Lymphocyte Reaction (MLR) Assay

- 35 This example shows that the polypeptides of the invention are active as a stimulator of the proliferation of stimulated T-lymphocytes. Compounds which stimulate proliferation of lymphocytes are useful therapeutically where enhancement of an immune response is beneficial. Compounds which inhibit proliferation of lymphocytes are useful therapeutically where suppression of an immune response is beneficial. A therapeutic agent may take the form of antagonists of the polypeptide of the invention, for  
40 example, murine-human chimeric, humanized or human antibodies against the polypeptide.



The basic protocol for this assay is described in Current Protocols in Immunology, unit 3.12; edited by J E Coligan, A M Kruisbeek, D H Marglies, E M Shevach, W Strober, National Institutes of Health, Published by John Wiley & Sons, Inc.

More specifically, in one assay variant, peripheral blood mononuclear cells (PBMC) are isolated from mammalian individuals, for example a human volunteer, by leukopheresis (one donor will supply stimulator PBMCs, the other donor will supply responder PBMCs). If desired, the cells are frozen in fetal bovine serum and DMSO after isolation. Frozen cells may be thawed overnight in assay media (37 °C, 5% CO<sub>2</sub>) and then washed and resuspended to 3 x 10<sup>6</sup> cells/ml of assay media (RPMI; 10% fetal bovine serum, 1% penicillin/streptomycin, 1% glutamine, 1% HEPES, 1% non-essential amino acids, 1% pyruvate).

The stimulator PBMCs are prepared by irradiating the cells (about 3000 Rads). The assay is prepared by plating in triplicate wells a mixture of:

100 µl of test sample diluted to 1% or to 0.1%

50 µl of irradiated stimulator cells and

50 µl of responder PBMC cells.

100 microliters of cell culture media or 100 microliter of CD4-IgG is used as the control. The wells are then incubated at 37 °C, 5% CO<sub>2</sub> for 4 days. On day 5 each well is pulsed with tritiated thymidine (1.0 mCi/well; Amersham). After 6 hours the cells are washed 3 times and then the uptake of the label is evaluated.

In another variant of this assay, PBMCs are isolated from the spleens of Balb/c mice and C57B6 mice. The cells are teased from freshly harvested spleens in assay media (RPMI; 10% fetal bovine serum, 1% penicillin/streptomycin, 1% glutamine, 1% HEPES, 1% non-essential amino acids, 1% pyruvate) and the PBMCs are isolated by overlaying these cells over Lympholyte M (Organon Teknika), centrifuging at 2000 rpm for 20 minutes, collecting and washing the mononuclear cell layer in assay media and resuspending the cells to 1 x 10<sup>7</sup> cells/ml of assay media. The assay is then conducted as described above. The results of this assay for compounds of the invention are shown below. Positive increases over control are considered positive with increases of greater than or equal to 180% being preferred. However, any value greater than control indicates a stimulatory effect for the test protein.

Table

	<u>PRO</u>	<u>PRO Concentration</u>	<u>Percent Increase Over Control</u>
30	PRO245	0.1%	189.7
	"	0.1%	193.7
	"	1.0%	212.5
	"	1.0%	300.5
	PRO217	0.1%	74.5
35	"	1.0%	89.5
	"	0.99 nM	97.0
	"	9.9 nM	122.3
	"	0.25 nM	144.8
	"	2.5 nM	126.9
40	PRO301	50.0%	109.4
	"	70.0 nM	133.7

	"	700.0 nM	83.6
		0.1%	58.7
	PRO301 (cont.)	1.0%	127.7
	"	0.1%	181.7
5	"	1.0%	187.3
	"	0.1%	127.5
	"	1.0%	108.3
	PRO266	0.1%	136.4
	"	0.1%	139.2
10	"	1.0%	189.8
	"	1.0%	245.1
	PRO335	50.0%	91.0
	"	50.0%	103.8
	"	0.1%	130.0
15	"	1.0%	180.2
	PRO331	50.0%	155.5
	"	0.1%	169.3
	"	1.0%	128.1
	"	0.1%	129.3
20	"	1.0%	162.5
	PRO326	50.0%	91.0
	"	50.0%	103.8
	"	0.1%	130.0
	"	1.0%	180.2
25			

### EXAMPLE 3

#### Skin Vascular Permeability Assay

This assay shows that certain polypeptides of the invention stimulate an immune response and induce inflammation by inducing mononuclear cell, eosinophil and PMN infiltration at the site of injection of the animal. This skin vascular permeability assay is conducted as follows. Hairless guinea pigs weighing 350 grams or more are anesthetized with ketamine (75-80 mg/Kg) and 5 mg/Kg xylazine intramuscularly (IM). A sample of purified polypeptide of the invention or a conditioned media test sample is injected intradermally onto the backs of the test animals with 100 uL per injection site. It is possible to have about 10-30, preferably about 16-24, injection sites per animal. One mL of Evans blue dye (1% in physiologic buffered saline) is injected intracardially. Blemishes at the injection sites are then measured (mm diameter) at 1hr and 6 hr post injection. Animals were sacrificed at 6 hrs after injection. Each skin injection site was biopsied and fixed in formalin. The skins were then prepared for histopathologic evaluation. Each site was evaluated for inflammatory cell infiltration into the skin. Sites with visible inflammatory cell inflammation were scored as positive. Inflammatory cells can be neutrophilic, eosinophilic, monocytic or lymphocytic. The results of this test for compounds of the invention is shown below.

In the Table below, at least a minimal perivascular infiltrate at the injection site is scored as positive, no infiltrate at the site of injection is scored as negative.

Table

<u>PRO</u>	<u>Hours Post Injection</u>	<u>Infiltrate Designation</u>
PRO245	24 hr	positive
PRO217	24 hr	positive
5 PRO301	24 hr	positive
PRO266	24 hr	positive
PRO335	24 hr	positive
PRO331	24 hr	positive
PRO326	24 hr	positive

10

## EXAMPLE 4

Human Co-Stimulation Assay

In addition to the activation signal mediated by the T cell receptor, T cell activation requires a costimulatory signal. One costimulatory signal is generated by the interaction of B7 (CD3) with CD28. In this assay, 96 well plates are coated with CD3 with or without CD28 and then human peripheral blood lymphocytes followed by a test protein, are added. Proliferation of the lymphocytes is determined by tritiated thymidine uptake. A positive assay indicates that the test protein provided a stimulatory signal for lymphocyte proliferation.

Material:

- 1) Hyclone D-PBS without Calcium, Magnesium
- 20 2) Nunc 96 well certified plates #4-39454
- 3) Anti-human CD3 Amac 0178 200 µg/ml stock
- 4) Anti-human CD28 Biodesign P42235M
- 5) Media: Gibco RPMI 1640 + 10 % Interger #1020-90 FBS, 1% Glu, 1% P/S, 50 µg/ml Gentamycin, 10 mM Hepes. Fresh for each assay.
- 25 6) Tritiated Thymidine
- 7) Frozen human peripheral blood lymphocytes (PBL) collected via a leukaphoresis procedure

Plates are prepared by coating 96 well flat bottom plates with anti-CD3 antibody (Amac) or anti-CD28 antibody (Biodesign) or both in Hyclone D-PBS without calcium and magnesium. Anti-CD3 antibody is used at 10 ng/well (50 µl of 200 ng/ml) and anti-CD28 antibody at 25 ng/well (50 µl of 0.5 µg/ml) in 100 µl total volume.

PBLs are isolated from human donors using standard leukaphoresis methods. The cell preparations are frozen in 50% fetal bovine serum and 50% DMSO until the assay is conducted. Cells are prepared by thawing and washing PBLs in media, resuspending PBLs in 25 mls of media and incubating at 37°C, 5% CO<sub>2</sub> overnight.

35 In the assay procedure, the coated plate is washed twice with PBS and the PBLs are washed and resuspended to  $1 \times 10^6$  cells/ml using 100 µL /well. 100 ul of a test protein or control media are added to the plate making a total volume per well of 200 µL. The plate is incubated for 72 hours. The plate is then pulsed for 6 hours with tritiated thymidine (1 mCi/well; Amersham) and the PBLs are harvested from the plates and evaluated for uptake of the tritiated thymidine.

## EXAMPLE 5

In situ Hybridization

*In situ* hybridization is a powerful and versatile technique for the detection and localization of nucleic acid sequences within cell or tissue preparations. It may be useful, for example, to identify sites of gene expression, analyze the tissue distribution of transcription, identify and localize viral infection, follow changes in specific mRNA synthesis and aid in chromosome mapping.

*In situ* hybridization was performed following an optimized version of the protocol by Lu and Gillett, Cell Vision 1: 169-176 (1994), using PCR-generated <sup>33</sup>P-labeled riboprobes. Briefly, formalin-fixed, paraffin-embedded human tissues were sectioned, deparaffinized, deproteinized in proteinase K (20 g/ml) for 15 minutes at 37°C, and further processed for *in situ* hybridization as described by Lu and Gillett, *supra*. A [<sup>33</sup>P] UTP-labeled antisense riboprobe was generated from a PCR product and hybridized at 55°C overnight. The slides were dipped in Kodak NTB2 nuclear track emulsion and exposed for 4 weeks.

<sup>33</sup>P-Riboprobe synthesis

6.0 µl (125 mCi) of <sup>33</sup>P-UTP (Amersham BF 1002, SA <2000 Ci/mmol) were speed vac dried.

To each tube containing dried <sup>33</sup>P-UTP, the following ingredients were added:

- 2.0 µl 5x transcription buffer
- 1.0 µl DTT (100 mM)
- 2.0 µl NTP mix (2.5 mM : 10 µl; each of 10 mM GTP, CTP & ATP + 10 µl H<sub>2</sub>O)
- 1.0 µl UTP (50 µM)
- 1.0 µl Rnasin
- 1.0 µl DNA template (1 µg)
- 1.0 µl H<sub>2</sub>O

The tubes were incubated at 37°C for one hour. 1.0 µL RQ1 DNase were added, followed by incubation at 37°C for 15 minutes. 90 µL TE (10 mM Tris pH 7.6/1mM EDTA pH 8.0) were added, and the mixture was pipetted onto DE81 paper. The remaining solution was loaded in a Microcon-50 ultrafiltration unit, and spun using program 10 (6 minutes). The filtration unit was inverted over a second tube and spun using program 2 (3 minutes). After the final recovery spin, 100 µL TE were added. 1 µL of the final product was pipetted on DE81 paper and counted in 6 ml of Biofluor II.

The probe was run on a TBE/urea gel. 1-3 µL of the probe or 5 µL of RNA Mrk III were added to 30 µL of loading buffer. After heating on a 95°C heat block for three minutes, the gel was immediately placed on ice. The wells of gel were flushed, the sample loaded, and run at 180-250 volts for 45 minutes. The gel was wrapped in saran wrap and exposed to XAR film with an intensifying screen in -70°C freezer one hour to overnight.

<sup>33</sup>P-Hybridization

*Pretreatment of frozen sections* The slides were removed from the freezer, placed on aluminium trays and thawed at room temperature for 5 minutes. The trays were placed in 55°C incubator for five minutes to reduce condensation. The slides were fixed for 10 minutes in 4% paraformaldehyde on ice in the fume hood, and washed in 0.5 x SSC for 5 minutes, at room temperature (25 ml 20 x SSC + 975 ml SQ H<sub>2</sub>O). After deproteinization in 0.5 µg/ml proteinase K for 10 minutes at 37°C (12.5µL of 10 mg/ml stock in 250 ml

prewarmed RNase-free RNase buffer), the sections were washed in 0.5 x SSC for 10 minutes at room temperature. The sections were dehydrated in 70%, 95%, 100% ethanol, 2 minutes each.

*Pretreatment of paraffin-embedded sections* The slides were deparaffinized, placed in SQ H<sub>2</sub>O, and rinsed twice in 2 x SSC at room temperature, for 5 minutes each time. The sections were deproteinated in 20 µg/ml proteinase K (500 µL of 10 mg/ml in 250 ml RNase-free RNase buffer; 37°C, 15 minutes) - human embryo, or 8 x proteinase K (100 µL in 250 ml RNase buffer, 37°C, 30 minutes) - formalin tissues. Subsequent rinsing in 0.5 x SSC and dehydration were performed as described above.

*Prehybridization* The slides were laid out in plastic box lined with Box buffer (4 x SSC, 50% formamide) - saturated filter paper. The tissue was covered with 50 µL of hybridization buffer (3.75g Dextran Sulfate + 6 ml SQ H<sub>2</sub>O), vortexed and heated in the microwave for 2 minutes with the cap loosened. After cooling on ice, 18.75 ml formamide, 3.75 ml 20 x SSC and 9 ml SQ H<sub>2</sub>O were added, the tissue was vortexed well, and incubated at 42°C for 1-4 hours.

*Hybridization* 1.0 x 10<sup>6</sup> cpm probe and 1.0 µL tRNA (50 mg/ml stock) per slide were heated at 95°C for 3 minutes. The slides were cooled on ice, and 48 µL hybridization buffer were added per slide. After vortexing, 50 µL <sup>33</sup>P mix were added to 50 µL prehybridization on slide. The slides were incubated overnight at 55°C.

*Washes* Washing was done 2x10 minutes with 2xSSC, EDTA at room temperature (400 ml 20 x SSC + 16 ml 0.25M EDTA, V<sub>f</sub>=4L), followed by RNaseA treatment at 37°C for 30 minutes (500 µL of 10 mg/ml in 250 ml RNase buffer = 20 µg/ml). The slides were washed 2x10 minutes with 2 x SSC, EDTA at room temperature. The stringency wash conditions were as follows: 2 hours at 55°C, 0.1 x SSC, EDTA (20 ml 20 x SSC + 16 ml EDTA, V<sub>f</sub>=4L).

#### DNA 35638 (1 TM receptor)

Expression was observed in the endothelium lining of a subset of fetal and placental vessels.

#### Endothelial

expression was confined to these tissue blocks. Expression was also observed over intermediate trophoblast cells of placenta.

Oligo C-120N: (SEQ ID NO:311)

GGA TTC TAA TAC GAC TCA CTA TAG GGC TGC GGC GGC TCA GGT CTT CAG TT

Oligo c-120P (SEQ ID NO:312)

CTA TGA AAT TAA CCC TCA CTA AAG GGA GCA TGG GAT GGG GAG GGA TAC GG

#### DNA 33094 (EGF Homolog)

A highly distinctive expression pattern was observed. In the human embryo expression was observed in outer smooth muscle layer of the GI tract, respiratory cartilage, branching respiratory epithelium, osteoblasts, tendons, gonad, in the optic nerve head and developing dermis. In the adult, expression was observed in the epidermal pegs of the chimp tongue, the basal epithelial / myoepithelial cells of the prostate and urinary bladder. Expression was also found in the alveolar lining cells of the adult lung, mesenchymal cells juxtaposed to erectile tissue in the penis and the cerebral cortex (probably glial cells). In the kidney, expression was only seen in disease, in cells surrounding thyroidized renal tubules.

Oligo D-200V (SEQ ID NO:313)

CTA TGA AAT TAA CCC TCA CTA AAG GGA ATA GCA GGC CAT CCC AGG ACA

Oligo D-200Z (SEQ ID NO:314)

5 CTA TGA AAT TAA CCC TCA CTA AAG GGA TGT CTT CCA TGC CAA CCT TC

#### EXAMPLE 6

##### Use of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 as a hybridization probe.

10 DNA comprising the coding sequence of full-length or mature PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 (as shown in Figure 1, SEQ ID NO:1; Figure 5C, SEQ ID NO:15; Figure 15, SEQ ID NO:75; Figures 20A-B, SEQ ID NO:237; Figures 25A-B, SEQ ID NO:262; Figure 32, SEQ ID NO:280; or Figures 37A-C, SEQ ID NO:299) is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO245, PRO217, PRO301, PRO266, 15 PRO335, PRO331 or PRO326) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 20 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be identified using standard techniques known in the art.

25

#### EXAMPLE 7

##### Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in *E. coli*

This example illustrates preparation of an unglycosylated form of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 by recombinant expression in *E. coli*.

The DNA sequence encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 30 is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., *Gene*, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The 35 vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in 40 Sambrook et al., *supra*. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell  
5 pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 protein can then be purified using a metal chelating column under conditions that allow tight binding of the protein.

PRO245, PRO217, PRO301 and PRO266 were expressed in *E. coli* in a poly-His tagged form, using the following procedure. The DNA encoding PRO245, PRO217, PRO301 and PRO266 was initially  
10 amplified using selected PCR primers. The primers contained restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector, and other useful sequences providing for efficient and reliable translation initiation, rapid purification on a metal chelation column, and proteolytic removal with enterokinase. The PCR-amplified, poly-His tagged sequences were then ligated into an expression vector, which was used to transform an *E. coli* host based on strain 52 (W3110 fuhA(tonA) lon  
15 galE rpoHts(htpRts) clpP(lacIq). Transformants were first grown in LB containing 50 mg/ml carbenicillin at 30°C with shaking until an O.D.600 of 3-5 was reached. Cultures were then diluted 50-100 fold into CRAP media (prepared by mixing 3.57 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.71 g sodium citrate-2H<sub>2</sub>O, 1.07 g KCl, 5.36 g Difco yeast extract, 5.36 g Sheffield hycase SF in 500 mL water, as well as 110 mM MPOS, pH 7.3, 0.55% (w/v) glucose and 7 mM MgSO<sub>4</sub>) and grown for approximately 20-30 hours at 30°C with shaking. Samples were removed  
20 to verify expression by SDS-PAGE analysis, and the bulk culture is centrifuged to pellet the cells. Cell pellets were frozen until purification and refolding.

*E. coli* paste from 0.5 to 1 L fermentations (6-10 g pellets) was resuspended in 10 volumes (w/v) in 7 M guanidine, 20 mM Tris, pH 8 buffer. Solid sodium sulfite and sodium tetrathionate is added to make final concentrations of 0.1M and 0.02 M, respectively, and the solution was stirred overnight at 4°C. This step  
25 results in a denatured protein with all cysteine residues blocked by sulfitolization. The solution was centrifuged at 40,000 rpm in a Beckman Ultracentrifuge for 30 min. The supernatant was diluted with 3-5 volumes of metal chelate column buffer (6 M guanidine, 20 mM Tris, pH 7.4) and filtered through 0.22 micron filters to clarify. Depending the clarified extract was loaded onto a 5 ml Qiagen Ni-NTA metal chelate column equilibrated in the metal chelate column buffer. The column was washed with additional  
30 buffer containing 50 mM imidazole (Calbiochem, Utrol grade), pH 7.4. The protein was eluted with buffer containing 250 mM imidazole. Fractions containing the desired protein were pooled and stored at 4°C. Protein concentration was estimated by its absorbance at 280 nm using the calculated extinction coefficient based on its amino acid sequence.

The proteins were refolded by diluting sample slowly into freshly prepared refolding buffer  
35 consisting of: 20 mM Tris, pH 8.6, 0.3 M NaCl, 2.5 M urea, 5 mM cysteine, 20 mM glycine and 1 mM EDTA. Refolding volumes were chosen so that the final protein concentration was between 50 to 100 micrograms/ml. The refolding solution was stirred gently at 4°C for 12-36 hours. The refolding reaction was quenched by the addition of TFA to a final concentration of 0.4% (pH of approximately 3). Before further purification of the protein, the solution was filtered through a 0.22 micron filter and acetonitrile was added to  
40 2-10% final concentration. The refolded protein was chromatographed on a Poros R1/H reversed phase column using a mobile buffer of 0.1% TFA with elution with a gradient of acetonitrile from 10 to 80%.

Aliquots of fractions with A280 absorbance were analyzed on SDS polyacrylamide gels and fractions containing homogeneous refolded protein were pooled. Generally, the properly refolded species of most proteins are eluted at the lowest concentrations of acetonitrile since those species are the most compact with their hydrophobic interiors shielded from interaction with the reversed phase resin. Aggregated species are usually eluted at higher acetonitrile concentrations. In addition to resolving misfolded forms of proteins from the desired form, the reversed phase step also removes endotoxin from the samples.

Fractions containing the desired folded PRO245, PRO217, PRO301 and PRO266 proteins, respectively, were pooled and the acetonitrile removed using a gentle stream of nitrogen directed at the solution. Proteins were formulated into 20 mM Hepes, pH 6.8 with 0.14 M sodium chloride and 4% mannitol by dialysis or by gel filtration using G25 Superfine (Pharmacia) resins equilibrated in the formulation buffer and sterile filtered.

#### EXAMPLE 8

##### Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in mammalian cells

This example illustrates preparation of a potentially glycosylated form of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 DNA using ligation methods such as described in Sambrook et al., *supra*. The resulting vector is called pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 ug pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 DNA is mixed with about 1 ug DNA encoding the VA RNA gene [Thimmappaya et al., *Cell*, 31:543 (1982)] and dissolved in 500 uL of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M CaCl<sub>2</sub>. To this mixture is added, dropwise, 500 uL of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM NaPO<sub>4</sub>, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 uCi/ml <sup>35</sup>S-cysteine and 200 uCi/ml <sup>35</sup>S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may be introduced into 293 cells transiently using the dextran sulfate method described by Sompariyac et al., *Proc. Natl. Acad. Sci.*, 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 ug pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran



precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 ug/ml bovine insulin and 0.1 ug/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be expressed in CHO cells. The pRK5-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be transfected into CHO cells using known reagents such as CaPO<sub>4</sub> or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as <sup>35</sup>S-methionine. After determining the presence of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be concentrated and purified by any selected method.

Epitope-tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may also be expressed in host CHO cells. The PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be concentrated and purified by any selected method, such as by Ni<sup>2+</sup>-chelate affinity chromatography.

PRO245, PRO217 and PRO301 were expressed in CHO cells by both a transient and a stable expression procedure.

Stable expression in CHO cells was performed using the following procedure. The proteins were expressed as an IgG construct (immunoadhesin), in which the coding sequences for the soluble forms (e.g. extracellular domains) of the respective proteins were fused to an IgG1 constant region sequence containing the hinge, CH2 and CH2 domains and/or is a poly-His tagged form.

Following PCR amplification, the respective DNAs were subcloned in a CHO expression vector using standard techniques as described in Ausubel et al., *Current Protocols of Molecular Biology*, Unit 3.16, John Wiley and Sons (1997). CHO expression vectors are constructed to have compatible restriction sites 5' and 3' of the DNA of interest to allow the convenient shuttling of cDNAs. The vector used expression in CHO cells is as described in Lucas et al., *Nucl. Acids Res.* 24: 9 (1774-1779 (1996), and uses the SV40 early promoter/enhancer to drive expression of the cDNA of interest and dihydrofolate reductase (DHFR). DHFR expression permits selection for stable maintenance of the plasmid following transfection.

Twelve micrograms of the desired plasmid DNA were introduced into approximately 10 million CHO cells using commercially available transfection reagents Superfect<sup>®</sup> (Quiagen), Dosper<sup>®</sup> or Fugene<sup>®</sup> (Boehringer Mannheim). The cells were grown and described in Lucas et al., supra. Approximately 3 x 10<sup>7</sup> cells are frozen in an ampule for further growth and production as described below.

The ampules containing the plasmid DNA were thawed by placement into water bath and mixed by vortexing. The contents were pipetted into a centrifuge tube containing 10 mLs of media and centrifuged at 1000 rpm for 5 minutes. The supernatant was aspirated and the cells were resuspended in 10 mL of selective media (0.2  $\mu$ m filtered PS20 with 5% 0.2  $\mu$ m diafiltered fetal bovine serum). The cells were then aliquoted  
5 into a 100 mL spinner containing 90 mL of selective media. After 1-2 days, the cells were transferred into a 250 mL spinner filled with 150 mL selective growth medium and incubated at 37°C. After another 2-3 days, a 250 mL, 500 mL and 2000 mL spinners were seeded with  $3 \times 10^5$  cells/mL. The cell media was exchanged with fresh media by centrifugation and resuspension in production medium. Although any suitable CHO media may be employed, a production medium described in US Patent No. 5,122,469, issued June 16, 1992  
10 was actually used. 3L production spinner is seeded at  $1.2 \times 10^6$  cells/mL. On day 0, the cell number pH were determined. On day 1, the spinner was sampled and sparging with filtered air was commenced. On day 2, the spinner was sampled, the temperature shifted to 33°C, and 30 mL of 500 g/L glucose and 0.6 mL of 10% antifoam (e.g., 35% polydimethylsiloxane emulsion, Dow Corning 365 Medical Grade Emulsion). Throughout the production, pH was adjusted as necessary to keep at around 7.2. After 10 days, or until  
15 viability dropped below 70%, the cell culture was harvested by centrifugation and filtering through a 0.22  $\mu$ m filter. The filtrate was either stored at 4°C or immediately loaded onto columns for purification.

For the poly-His tagged constructs, the proteins were purified using a Ni-NTA column (Qiagen). Before purification, imidazole was added to the conditioned media to a concentration of 5 mM. The conditioned media was pumped onto a 6 ml Ni-NTA column equilibrated in 20 mM Hepes, pH 7.4, buffer  
20 containing 0.3 M NaCl and 5 mM imidazole at a flow rate of 4-5 ml/min. at 4°C. After loading, the column was washed with additional equilibration buffer and the protein eluted with equilibration buffer containing 0.25 M imidazole. The highly purified protein was subsequently desalted into a storage buffer containing 10 mM Hepes, 0.14 M NaCl and 4% mannitol, pH 6.8, with a 25 ml G25 Superfine (Pharmacia) column and stored at -80°C.

Immunoadhesin (Fc containing) constructs of were purified from the conditioned media as follows. The conditioned medium was pumped onto a 5 ml Protein A column (Pharmacia) which had been equilibrated in 20 mM Na phosphate buffer, pH 6.8. After loading, the column was washed extensively with equilibration buffer before elution with 100 mM citric acid, pH 3.5. The eluted protein was immediately neutralized by collecting 1 ml fractions into tubes containing 275  $\mu$ L of 1 M Tris buffer, pH 9. The highly purified protein  
30 was subsequently desalted into storage buffer as described above for the poly-His tagged proteins. The homogeneity was assessed by SDS polyacrylamide gels and by N-terminal amino acid sequencing by Edman degradation.

PRO326 was also produced by transient expression in COS cells.

#### EXAMPLE 9

##### 35 Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in Yeast

The following method describes recombinant expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 from the ADH2/GAPDH promoter. DNA  
40 encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of PRO245,

PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326. For secretion, DNA encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, a native PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 signal peptide or other mammalian signal peptide, or, for example, a yeast alpha-factor or invertase secretory signal/leader sequence, and linker sequences (if needed) for expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 may further be purified using selected column chromatography resins.

#### EXAMPLE 10

##### Expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in Baculovirus-Infected Insect Cells

The following method describes recombinant expression of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 in Baculovirus-infected insect cells.

The sequence coding for PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the sequence encoding PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 or the desired portion of the coding sequence of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 [such as the sequence encoding the extracellular domain of a transmembrane protein or the sequence encoding the mature protein if the protein is extracellular] is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (PharMingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 - 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression are performed as described by O'Reilley et al., Baculovirus expression vectors: A Laboratory Manual, Oxford: Oxford University Press (1994).

Expressed poly-his tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can then be purified, for example, by Ni<sup>2+</sup>-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM MgCl<sub>2</sub>; 0.1 mM EDTA; 10% glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl,

10% glycerol, pH 7.8) and filtered through a 0.45  $\mu$ m filter. A  $\text{Ni}^{2+}$ -NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline  $A_{280}$  with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching  $A_{280}$  baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or Western blot with  $\text{Ni}^{2+}$ -NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His<sub>10</sub>-tagged PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

PRO245, PRO301 and PRO266 were expressed in baculovirus infected Sf9 insect cells. While the expression was actually performed in a 0.5-2 L scale, it can be readily scaled up for larger (e.g. 8 L) preparations. The proteins were expressed as an IgG construct (immunoadhesin), in which the protein extracellular region was fused to an IgG1 constant region sequence containing the hinge, CH2 and CH3 domains and/or in poly-His tagged forms.

Following PCR amplification, the respective coding sequences were subcloned into a baculovirus expression vector (pb.PH.IgG for IgG fusions and pb.PH.His.c for poly-His tagged proteins), and the vector and Baculogold<sup>®</sup> baculovirus DNA (Pharmingen) were co-transfected into 105 *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711), using Lipofectin (Gibco BRL). pb.PH.IgG and pb.PH.His are modifications of the commercially available baculovirus expression vector pVL1393 (Pharmingen), with modified polylinker regions to include the His or Fc tag sequences. The cells were grown in Hink's TNM-FH medium supplemented with 10% FBS (Hyclone). Cells were incubated for 5 days at 28°C. The supernatant was harvested and subsequently used for the first viral amplification by infecting Sf9 cells in Hink's TNM-FH medium supplemented with 10% FBS at an approximate multiplicity of infection (MOI) of 10. Cells were incubated for 3 days at 28°C. The supernatant was harvested and the expression of the constructs in the baculovirus expression vector was determined by batch binding of 1 ml of supernatant to 25 mL of Ni-NTA beads (QIAGEN) for histidine tagged proteins or Protein-A Sepharose CL-4B beads (Pharmacia) for IgG tagged proteins followed by SDS-PAGE analysis comparing to a known concentration of protein standard by Coomassie blue staining.

The first viral amplification supernatant was used to infect a spinner culture (500 ml) of Sf9 cells grown in ESF-921 medium (Expression Systems LLC) at an approximate MOI of 0.1. Cells were incubated for 3 days at 28°C. The supernatant was harvested and filtered. Batch binding and SDS-PAGE analysis was repeated, as necessary, until expression of the spinner culture was confirmed.

The conditioned medium from the transfected cells (0.5 to 3 L) was harvested by centrifugation to remove the cells and filtered through 0.22 micron filters. For the poly-His tagged constructs, the protein construct were purified using a Ni-NTA column (Qiagen). Before purification, imidazole was added to the conditioned media to a concentration of 5 mM. The conditioned media were pumped onto a 6 ml Ni-NTA column equilibrated in 20 mM Hepes, pH 7.4, buffer containing 0.3 M NaCl and 5 mM imidazole at a flow

rate of 4-5 ml/min. at 4°C. After loading, the column was washed with additional equilibration buffer and the protein eluted with equilibration buffer containing 0.25 M imidazole. The highly purified protein was subsequently desalted into a storage buffer containing 10 mM Hepes, 0.14 M NaCl and 4% mannitol, pH 6.8, with a 25 ml G25 Superfine (Pharmacia) column and stored at -80°C.

- 5 Immunoadhesin (Fc containing) constructs of proteins were purified from the conditioned media as follows. The conditioned media were pumped onto a 5 ml Protein A column (Pharmacia) which had been equilibrated in 20 mM Na phosphate buffer, pH 6.8. After loading, the column was washed extensively with equilibration buffer before elution with 100 mM citric acid, pH 3.5. The eluted protein was immediately neutralized by collecting 1 ml fractions into tubes containing 275 mL of 1 M Tris buffer, pH 9. The highly  
10 purified protein was subsequently desalted into storage buffer as described above for the poly-His tagged proteins. The homogeneity of the proteins was verified by SDS polyacrylamide gel (PAGE) electrophoresis and N-terminal amino acid sequencing by Edman degradation.

PRO245, PRO217, PRO301, PRO266, PRO331 and PRO326 were also expressed in baculovirus infected High-5 cells using an analogous procedure.

15 **EXAMPLE 11**

Preparation of Antibodies that Bind PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326

This example illustrates preparation of monoclonal antibodies which can specifically bind PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326.

- Techniques for producing the monoclonal antibodies are known in the art and are described, for  
20 instance, in Goding, *supra*. Immunogens that may be employed include purified PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326, fusion proteins containing PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326, and cells expressing recombinant PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

- 25 Mice, such as Balb/c, are immunized with the PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the  
30 selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326.

- 35 Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

- 40 The hybridoma cells will be screened in an ELISA for reactivity against PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326. Determination of "positive" hybridoma cells secreting the desired

monoclonal antibodies against PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO245, PRO217, PRO301, PRO266, PRO335, PRO331 or PRO326 monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

#### Deposit of Material

10 The following materials have been deposited with the American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209, USA (ATCC):

<u>Material</u>	<u>ATCC Dep. No.</u>	<u>Deposit Date</u>
DNA40981	209439	7 November 1997
DNA37140	209489	21 November 1997
15 DNA41388	209927	2 June 1998
DNA35638	209265	17 September 1997
DNA37150	209401	17 October 1997
DNA33094	209256	16 September 1997
DNA32292	209258	16 September 1997
20 DNA32279	209259	16 September 1997
DNA40628	209432	7 November 1997

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC § 122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of

the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

## Claims:

1. A composition, comprising a PRO245 polypeptide, agonist or fragment thereof and a carrier or excipient, useful for:
  - (a) increasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - 5 (b) stimulating or enhancing an immune response in a mammal in need thereof, or
  - (c) increasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
2. Use of a PRO245 polypeptide, agonist or a fragment thereof to prepare a composition useful for:
  - 10 (a) increasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - (b) stimulating or enhancing an immune response in a mammal in need thereof, or
  - (c) increasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
3. A composition, comprising a PRO245 polypeptide, antagonist or a fragment thereof and a carrier or excipient, useful for:
  - 15 (a) decreasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - (b) inhibiting or reducing an immune response in a mammal in need thereof, or
  - (c) decreasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
- 20 4. Use of a PRO245 polypeptide, antagonist or a fragment thereof to prepare a composition useful for:
  - (a) decreasing infiltration of inflammatory cells into a tissue of a mammal in need thereof,
  - (b) inhibiting or reducing an immune response in a mammal in need thereof, or
  - (c) decreasing the proliferation of T-lymphocytes in a mammal in need thereof in response to an antigen.
- 25 5. A method of treating an immune related disorder, such as a T cell mediated disorder, in a mammal in need thereof, comprising administering to the mammal an effective amount of a PRO245 polypeptide, an agonist antibody thereof, an antagonist antibody thereto, or a fragment thereof.
6. The method of claim 5, wherein the disorder is selected from systemic lupus erythematosus,
  - 30 rheumatoid arthritis, juvenile chronic arthritis, spondyloarthropathies, systemic sclerosis (scleroderma), idiopathic inflammatory myopathies (dermatomyositis, polymyositis), Sjsgren's syndrome, systemic vasculitis, sarcoidosis, autoimmune hemolytic anemia (immune pancytopenia, paroxysmal nocturnal hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic purpura, immune-mediated thrombocytopenia), thyroiditis (Grave's disease, Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis,
  - 35 atrophic thyroiditis), diabetes mellitus, immune-mediated renal disease (glomerulonephritis, tubulointerstitial nephritis), demyelinating diseases of the central and peripheral nervous systems such as multiple sclerosis, idiopathic demyelinating polyneuropathy or Guillain-Barré syndrome, and chronic inflammatory demyelinating polyneuropathy, hepatobiliary diseases such as infectious hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune chronic active hepatitis, primary biliary cirrhosis,
  - 40 granulomatous hepatitis, and sclerosing cholangitis, inflammatory and fibrotic lung diseases such as inflammatory bowel disease (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and Whipple's disease. autoimmune or immune-mediated skin diseases including bullous skin diseases, erythema multiforme



and contact dermatitis, psoriasis, allergic diseases such as asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity and urticaria, immunologic diseases of the lung such as eosinophilic pneumonias, idiopathic pulmonary fibrosis and hypersensitivity pneumonitis, transplantation associated diseases including graft rejection and graft -versus-host-disease.

5           7.       The composition or use of any of the preceding claims, wherein the antibody is a monoclonal antibody.

          8.       The composition or use of any of the preceding claims, wherein the antibody is an antibody fragment or a single-chain antibody.

          9.       The composition or use of any of the preceding claims, wherein the antibody has nonhuman  
10 complementarity determining region (CDR) residues and human framework region (FR) residues.

          10.      A method for determining the presence of a PRO245 polypeptide, comprising exposing a cell suspected of containing the PRO245 polypeptide to an anti-PRO245 antibody and determining binding of the antibody to the cell.

          11.      A method of diagnosing an immune related disease in a mammal, comprising detecting the  
15 level of expression of a gene encoding a PRO245 polypeptide (a) in a test sample of tissue cells obtained from the mammal, and (b) in a control sample of known normal tissue cells of the same cell type, wherein a higher expression level in the test sample indicates the presence of immune related disease in the mammal from which the test tissue cells were obtained.

          12.      A method of diagnosing an immune related disease in a mammal, comprising (a) contacting  
20 an anti-PRO245 antibody with a test sample of tissue cells obtained from the mammal, and (b) detecting the formation of a complex between the antibody and the polypeptide in the test sample.

          13.      An immune related disease diagnostic kit, comprising an anti-PRO245 antibody or fragment thereof and a carrier in suitable packaging.

          14.      The kit of claim 13, further comprising instructions for using the antibody to detect a  
25 PRO245 polypeptide.

          15.      An article of manufacture, comprising:  
                  a container;  
                  a label on the container; and  
                  a composition comprising an active agent contained within the container; wherein the  
30 composition is effective for stimulating or enhancing an immune response in a mammal, the label on the container indicates that the composition can be used for treating an immune related disease, and the active agent in the composition is an agent inhibiting the expression and/or activity of a PRO245 polypeptide.

          16.      The article of manufacture of claim 21 wherein said active agent is an anti-PRO245 antibody.

35           17.      A method for identifying a compound capable of inhibiting the expression or activity of a PRO245 polypeptide, comprising contacting a candidate compound with a PRO245 polypeptide under conditions and for a time sufficient to allow these two components to interact.

          18.      The method of claim 17, wherein the candidate compound or the PRO245 polypeptide is immobilized on a solid support.

40           19.      The method of claim 18, wherein the non-immobilized component carries a detectable label.

1 / 108

## SEQ ID NO:1

CCCAGAAGTTCAAGGGCCCCCGGCCTCCTGCGCTCCTGCCGCCGGGACCCCTCGACCTCCT  
CAGAGCAGCCGGCTGCCGCCCCGGGAAGATGGCGAGGAGGAGCCGCCACCGCCTCCTCCT  
GCTGCTGCTGCGCTACCTGGTGGTCGCCCTGGGCTATCATAAGGCCTATGGGTTTTCTGC  
CCCAAAGACCAACAAGTAGTCACAGCAGTAGAGTACCAAGAGGCTATTTTAGCCTGCAA  
AACCCCAAAGAAGACTGTTTCCTCCAGATTAGAGTGGAAGAACTGGGTTCGGAGTGTCTC  
CTTTGTCTACTATCAACAGACTCTTCAAGGTGATTTTAAAAATCGAGCTGAGATGATAGA  
TTTCAATATCCGGATCAAAAATGTGACAAGAAGTGATGCGGGGAAATATCGTTGTGAAGT  
TAGTGCCCCATCTGAGCAAGGCCAAAACCTGGAAGAGGATACAGTCACTCTGGAAGTATT  
AGTGGCTCCAGCAGTTCCATCATGTGAAGTACCCTCTTCTGCTCTGAGTGGAAGTGTGGT  
AGAGCTACGATGTCAAGACAAAGAAGGGAATCCAGCTCCTGAATACACATGGTTTAAGGA  
TGGCATCCGTTTGCTAGAAAATCCCAGACTTGGCTCCCAAAGCACCAACAGCTCATACAC  
AATGAATACAAAACTGGAAGTCTGCAATTTAATACTGTTTCCAACTGGACACTGGAGA  
ATATTCCCTGTGAAGCCCGCAATTCTGTTGGATATCGCAGGTGTCCTGGGAAACGAATGCA  
AGTAGATGATCTCAACATAAGTGGCATCATAGCAGCCGTAGTAGTTGTGGCCTTAGTGAT  
TTCCGTTTGTGGCCTTGGTGTATGCTATGCTCAGAGGAAAGGCTACTTTTCAAAGAAAC  
CTCCTTCCAGAAGAGTAATTCTTCATCTAAAGCCACGACAATGAGTGAAAATGTGCAGTG  
GCTCACGCCTGTAATCCCAGCACTTTGGAAGGCCGCGGCGGGCGGATCACGAGGTCAGGA  
GTTCTAGACCAGTCTGGCCAATATGGTGAAACCCCATCTCTACTAAAATACAAAATTAG  
CTGGGCATGGTGGCATGTGCCTGCAGTTCCAGCTGCTTGGGAGACAGGAGAATCACTTGA  
ACCCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCAGTCCAGCCTGGGTAA  
CAGAGCAAGATTCCATCTCAAAAAATAAAATAAATAAATAAATAAATAAATACTGGTTTTTACC  
TGTAAGATTCTTACAATAAATATAGCTTGATATTC

FIG. 1

2 / 108

SEQ ID NO:2

MARRSRHRLLLLLLLRYLVVALGYHKAYGFSAPKDQQVVTAVEYQEAILACKTPKKTVSSR  
LEWKKLGRSVSFVYYQQTLQGDFKNRAEMIDFNIRIKNVTRSDAGKYRCEVSAPSEQGQN  
LEEDTVTLEVLVAPAVPSCCEVPSSALSGTVVELRCQDKEGNPAPEYTWFKDGIRLLENPR  
LGSQSTNSSYTMNTKTGTLOFNTVSKLDTGEYSCEARNVGYRRCPGKRMQVDDLNISGI  
IAAVVVVALVISVCGLGVCYAQRKGYFSKETSFQKSNSSSKATTMSENVQWLTPVIPALW  
KAAAGGSRGQEF

FIG. 2

3 / 108

SEQ ID NO:8	2715631	1	CTGGGTCGGAGTGTCTCCTTTGTCTACTATCAACAGACTCTTCAAGGTGA
	2715631	51	TTTTTAAATAATCGAGCTGAGATGATAGATTTCATATATCCGGATCAA
SEQ ID NO:9	2715631	96	AAATGTACAAGAAGTGTATGCGGGGAAATATCGTTGTGAAGTTAGTGCCC
	1622388	1	CTCGAGCCGCTCGAGCCGTGCGGGGAAATATCGTTGTGAAGTTAGTGCCC
SEQ ID NO:7	<DNA30954>	1	CTCGAGCCGCTCGAGCCGTGCGGGGAAATATCGTTGTGAAGTTAGTGCCC
	2715631	146	CATCTGAGCAAGGCCAAAACCTGGAAGAGGATACAGTCACTCTGGAAAGTA
	1622388	51	CATCTGAGCAAGGCCAAAACCTGGAAGAGGATACAGTCACTCTGGAAAGTA
	<DNA30954>	51	CATCTGAGCAAGGCCAAAACCTGGAAGAGGATACAGTCACTCTGGAAAGTA
SEQ ID NO:10	2715631	196	TTAGTGGCTCCAGCAGTTCCTCATCATGTGAAGTA
	1622388	101	TTAGTGGNTCCAGCAGNTCCATCATGTGAAGTACCTCTCTGCTCTGAG
	T89217	1	GCAGTTCCTCATCATGTGAAGTACCTCTCTGCTCTGAG
	<DNA30954>	101	TTAGTGGCTCCAGCAGTTCCTCATCATGTGAAGTACCTCTCTGCTCTGAG
SEQ ID NO:11	1622388	151	TGGAACCTGTGGTAGAGCTACGATGTCAAGACAAAGAGGGAATCCAGCTC
	T89217	39	TGGAACCTGTGGTAGAGCTACGATGTCAAGACAAAGAGGGAATCCAGCTC
	1861250	1	GGTAGAGCTACGATGTCAAGACAAAGAGGGAATCCAGCTC
	<DNA30954>	151	TGGAACCTGTGGTAGAGCTACGATGTCAAGACAAAGAGGGAATCCAGCTC
	1622388	201	CTGAATACACATGGTTTAAGGATGGCATCCGTTTGCTAGAA
	T89217	89	CTGAATACACATGGTTTAAGGATGGCATCCGTTTGCTAGAAATCCCAGA
	1861250	42	CTGAATACACATGGTTTAAGGATGGCATCCGTTTGCTAGAAATCCCAGA
	<DNA30954>	201	CTGAATACACATGGTTTAAGGATGGCATCCGTTTGCTAGAAATCCCAGA

FIG. 3A

4 / 108

T89217	139	CTTGGCTCCCAAAGCACCAACAGCTCATAACAATGAATACAAAACTGG
1861250	92	CTTGGCTCCCAAAGCACCAACAGCTCATAACAATGAATACAAAACTGG
<DNA30954>	251	CTTGGCTCCCAAAGCACCAACAGCTCATAACAATGAATACAAAACTGG
T89217	189	AACTCTGCAATTTAATACTGTTTCCAAACTGGACACTGGAGAATATTCCT
1861250	142	AACTCTGCAATTTAATACTGTTTCCAAACTGGACACTGGAGAATATTCCT
<DNA30954>	301	AACTCTGCAATTTAATACTGTTTCCAAACTGGACACTGGAGAATATTCCT
T89217	239	GTGAAGCCCGCAATTCTGTTGGATATCGCAGGTGTCCTGGGAAACGAAT
1861250	192	GTGAAGCCCGCAATTCTGTTGGATATCGCAGGTGTCCTGGG-AAACGAAT
<DNA30954>	351	GTGAAGCCCGCAATTCTGTTGGATATCGCAGGTGTCCTGGGAAACGAAT
T89217	289	GCAAGTAGATGAT
1861250	242	GCAAGTAGATGAT
<DNA30954>	401	GCAAGTAGATGAT

FIG. 3B

SEQ ID NO:12	DNA35638	1228 LFFEMESCVTQAGLQWRDLSSLQPPPGFK*FSCL--PSSWNCRHHMPPCPANFCILVEM ** . * . ** * *** ***** **** * **. ** ** **** *
SEQ ID NO:3	HSU22376_2	555 LFTQTSPVADPTGVQWHDFGSLQLPPPGFKRFSCLSLPRSWDYRHPPRPANFEFLVET
	DNA35638	1054 GFHHIGQTGLELLTS ** *. ** *****
	HSU22376_2	615 GFLHVGQAGLELLTS

**FIG. 4**

## ss.DNA 32279

SEQ ID NO.13

CTCGAGCCG AGCGGCGCG GGAAGGGCT CTCCTTCCAG CGCGGAGCAC TGGGCCCTGG CAGACGCCCC AAGATTGTTG TGAGGAGTCT AGCCAGTTGG 100  
 TGAGCGCTGT AATCTGAACC AGCTGTGTCC AGACTGAGGC CCCATTGCA TTGTTTAAACA TACTTAGAAA ATGAAGTGTT CATTTTAAAC ATTCTTCCTC 200  
 CAATTGGTTT AATGCTGAAT TACTGAAGAG GGCTAAGCAA AACCAGGTGC TTGCGCTGAG GGCTCTGCAG TGCGTGGGAG GACCCCGGCG CTCCTCCCGT 300  
 GTCTCTTCCA CGACTCGCTC GGCCCTCTG GAATAAACA CCCGCGAGCC CCGAGGGCCC AGAGGAGGCC GACGTGCCCG AGCTCCTCCG GGGGTCCCGC 400  
 CCGGAGCTT TCTTCTCGCC TTCGCATCTC CTCCTGCGC GTCTTGGACA TGCCAGGAAT AAAAAAGGATA CTCACTGTGA CCATTCTGGC TCTCTGTCTT 500  
 CCAAGCCCTG GGAATGCACA GGCACAGTGC ACGAATGGCT TTGACCTGGA TCGCCAGTCA GGACAGTGTT TAGATATTGA TGAATGCCGA ACCATCCCCG 600  
 AGGCCTGCCG AGGAGACATG ATGTGTGTTA ACCAAATGG CGGGTATTTA TGCATTCCCC GGACAAACCC TGTGTATCGA GGGCCCTACT CGAACCCCTA 700  
 CTCGACCCCC TACTCAGGTC CGTACCCAGC AGCTGCCCCA CCACTCTCAG CTCCAAAC TA TCCACGATC TCCAGGCCCTC TTATATGCCG CTTTGGATAC 800  
 CAGATGGATG AAAGCAACCA ATGTGTGGAT GTGGACGAGT GTGCAACAGA TTCCCACCAG TGCAACCCCA CCCAGATCTG CATCAATACT GAAGGCGGGT 900  
 ACACCTGCTC CTGCACCGAC GGATATTGGC TTCTGGAAGG CCAGTGCCTA GACATTGATG AATGTCGCTA TGGTTACTGC CAGCAGCTCT GTGCGAATGT 1000  
 TCCTGGATCC TATTCTTGTA CATGCAACCC TGGTTTTACC CTCAATGAGG ATGGAAGGTC TTGCCAAGAT GTGAACGAGT GTGCCACCGA GAACCCCTGC 1100  
 GTGCAAAACCT GCGTCAACAC CTACGGCTCT CTCATCTGCC GCTGTGACCC AGGATATGAA CTTGAGGAAG ATGGCGTTCA TTGCAGTGAT ATGGACGAGT 1200  
 GCAGCTTCTC TGAGTTCCCTC TGCCAACATG ASTGTGTGAA CCAGCCCGGC ACATACTTCT GCTCCTGCC TCCAGGCTAC ATCCTGCTGG ATGACAACCG 1300  
 AAGCTGCCAA GACATCAACG AATGTGAGCA CAGGAACCAAC ACGTGCAACC TGCAGCAGAC GTGCTACAAT TTACAAGGGG GCTTCAAATG CATCGACCCC 1400  
 ATCCGCTGTG AGGAGCCTTA TCTGAGGATC AGTGATAACC GCTGTATGTG TCCTGCTGAG AACCCCTGGCT GCAGAGACCA GCCCTTTACC ATCTTGATAC 1500  
 GGGACATGGA CGTGGTGTCA GGACGCTCCG TTCCCGCTGA CATCTTCCAA ATGCAAGCCA CGACCCGCTA CCCTGGGGCC TATTACATTT TCCAGATCAA 1600  
 ATCTGGGAAT GAGGGCAGAG AATTTTACAT GCGGCAACG GGGCCCATCA GTGCCACCTT GGTGATGACA CGCCCCATCA AAGGGCCCCG GGAAATCCAG 1700  
 CTGGACTTGG AAATGATCAC TGTCAACACT GTCATCAACT TCAGAGGCAG CTCGGTGATC CGACTGCCGA TATATGTGC GCAGTACCCA TTCTGAGCCT 1800

FIG. 5A-1

7 / 108

CTGCTGAACG TTTCCCCGAA GAGTCAGCCC CGACTTCCTG ACTCTCACCT GTACTATTGC AGACCTGTCA CCTGCGAGGA CTTGCCACCC CCAGTTCCTA 2000  
TGACACAGTT ATCAAAAAGT ATTATCATTG CTCCTCTGAT AGAAGATTGT TGGTGAATTT TCAAGGCCCT CAGTTTATTT CAAAGAAAAT 2100  
AGATTAGGTT TCGGGGGGTC TGAGTCTATG TTCAAAGACT GTGAACAGCT TGCTGTCACT TCTTCACCTC TTCCACTCCT TCTCTCACTG TGTTACTGCT 2200  
TTGCAAAGAC CCGGGGAGCTG GCGGGGAACC CTGGGAGTAG CTAGTTTGCT TTTTGGGTAC ACAGAGAAGG CTATGTAAC AAACCCACAGC AGGATCGAAG 2300  
GGTTTTTAGA GAATGTGTTT CAAAACCATG CCTGGTATTT TCAACCATAA AAGAAGTTTC AGTTGTCCCTT AAATTTGTAT AACGGTTTAA TTCTGTCTTG 2400  
TTCATTTTGA GTATTTTTAA AAAATATGTC GTAGAATTCC TTCGAAAGGC CTTACAGACAC ATGCTATGTT CTGTCTTCCC AAACCCAGTC TCCTCTCCAT 2500  
TTTAGCCCCAG TGTTTTCTTT GAGGACCCCT TAATCTTGCT TTCTTTAGAA TTTTACCCA ATGGATTGG AATGCAGAGG TCTCCAAACT GATTAAATAT 2600  
TTGAAGAGA 2609

FIG. 5A-2



## ss.DNA32292

SEQ ID NO:14

GCGCGGAGCA GCACGGCCGC AGGACCTGGA GCTCCGGCTG CGTCTTCCCG CAGCGCTACC CGCATGCGC CTGCCGCGCC GGGCCGCGCT GGGGCTCCTG 100  
 CCGCTTCTGC TGCTGCTGCC GCGCGCGCCG GAGGCCGCCA AGAAGCCGAC GCCCTGCCAC CGGTGCCGGG GGCTGGTGGA CAAGTTTAAC CAGGGGATGG 200  
 TGGACACCGC AAAGAAGAAC TTTGGCGGCG GGAACACGGC TTGGGAGGAA AAGACGCTGT CCAAGTACGA GTCCAGCGAG ATTGCGCTGC TGGAGATCCT 300  
 GGAGGGGCTG TGGGAGAGCA GCGACTTCGA ATGCAATCAG ATGCTAGAGG CGCAGGAGGA GCACCTGGAG GCCTGGTGGC TGCAGCTGAA GAGCGAATAT 400  
 CCTGACTTAT TCGAGTGGTT TTGTGTGAAG ACACGTAAAG TGTGCTGCTC TCCAGGAACC TACGGTCCCG ACTGTCTCGC ATGCCAGGGC GGATCCCAGA 500  
 GGCCTGCAG CCGGAATGGC CACTGCAGCG GAGATGGGAG CAGACAGGGC GACGGGTCTT GCCGTGCCA CATGGGGTAC CAGGGCCCGC TGTGCACTGA 600  
 CTGCATGGAC GGCTACTTCA GCTCGCTCCG GAACGAGACC CACAGCATCT GCACAGCCTG TGACGAGTCC TGCAAGACGT GCTCGGGCCT GACCAACAGA 700  
 GACTGCCGCG AGTGTGAAGT GGGCTGGGTG CTGGACGAGG GCGCCTGTGT GGATGTGGAC GAGTGTGCGG CCGAGCCGCC TCCCTGCAGC GCTGCGCAGT 800  
 TCTGTAAGAA CGCCAACGGC TCCTACACGT GCGAAGAGTG TGACTCCAGC TGTGTGGGT GCACAGGGGA AGGCCCAGGA AACTGTAAAG AGTGTATCTC 900  
 TGGCTACCGC AGGGAGCACG GACAGTGTGC AGATGTGGAC GAGTGCTCAC TAGCAGAAAA AACCTGTGTG AGGAAAAACG AAAACTGCTA CAATACTCCA 1000  
 GGGAGCTACG TCTGTGTGTG TCCTGACGGC TTGGAAGANA CGGAAGATGC CTGTGTGCGG CCGGCAGAGG CTGAAGCCAC AGAAGGAGAA AGCCCGACAC 1100  
 AGCTGCCCTC CCGCGAAGAC CTGTAATGTG CCGGACTTAC CCTTTAAATT ATTCAGAAGG ATGTCCCGTG GAAAAATGTG CCCTGAGGAT GCCGTCTCCT 1200  
 GCAGTGGACA GCGCGCGGGA GAGGCTGCCT GCTCTCTAAC GGTGTATTCT CATTTGTCCC TTAACAGACT GCATTTCTTG GTTGTCTTA AACAGACTTG 1300  
 TATATTTTGA TACAGTTCTT TGTAAATAAAA TTGACCAATTG TAGGTAATCA GGAGGAAAAA AAAA 1364

SUBSTITUTE SHEET (RULE 26)

FIG. 5B

## ss.DNA33094

SEQ ID NO:15

CCAGCGCGG AGCGACGCG CCCAGCCGTC TAAACGGGAA CAGCCCTGGC TGAGGGAGCT GCAGCGCAGC AGAGTATCTG ACGGCGCCAG GTTGCCTAGG 100  
 TGGCGCACGA GGAGTTTCC CGGCAGCGAG GAGGTCTCTGA GCAGCATGGC CCGGAGGAGC GCCTTCCCTG CCGCCGCGCT CTGGCTCTGG AGCATCCTCC 200  
 TGTGCTGCT GGCACCTGCG GCGAGGCGG GCGCGCGGCA GGAGGAGAGC CTGTACCTAT GGATCGATGC TCACCAGGCA AGAGTACTCA TAGGATTIGA 300  
 AGAAGATATC CTGATTGTTT CAGAGGGGAA AATGGCACCT TTACACATG ATTTACAGAA AGCGCAACAG AGAATGCCAG CTATTCTCTG CAATATCCAT 400  
 TCCATGAATT TTACCTGGCA AGCTGCAGG CAGGCAGAAT ACTTCTATGA ATTCTGTCC TTGCGCTCCC TGGATAAAGG CATCATGGCA GATCCAACCG 500  
 TCAATGTCCC TCTGCTGGGA ACAGTGCCCT ACAAGGCATC AGTTGTTCAA GTTGGTTTCC CATGTCTTGG AAAACAGGAT GGGGTGGCAG CATTGGAAGT 600  
 GGATGTGATT GTTATGAATT CTGAAGGCAA CACCATCTC CAAACACCTC AAAATGCTAT CTTCTTTAAA ACATGTCAAC ARGCTGAGTG CCCAGGCGGG 700  
 TGCCGAAATG GAGGCTTTTG TAATGAAAGA CGCATCTGCG AGTGCTCTGA TGGGTTCCAC GGACCTCACT GTGAGAAAGC CCTTTGTACC CCACGATGTA 800  
 TGAATGGTGG ACTTTGTGTG ACTCCTGGT TCTGCATCTG CCCACCTGGA TTCTATGGAG TGAAGTGTGA CAAAGCAAAC TGCTCAACCA CCTGCTTTAA 900  
 TGGAGGGACC TGTTTCTACC CTGGAAAATG TATTGGCCCT CCAGGACTAG AGGGAGAGCA GTGTGAAATC AGCAAATGCC CACAACCCCTG TCGAAATGGA 1000  
 GGTAATGCA TTGGTAAAG CAATGTAAAG TGTTCCAAAG GTTACCAGGG AGACCTCTGT TCAAAGCCCTG TCTGCGAGCC TGGCTGTGGT GCACATGGAA 1100  
 CCTGCCATGA ACCCAACAA TGCCAATGTC AAGAAGGTTG GCATGGAAGA CACTGCAATA AAAGGTACGA AGCCAGCCTC ATACATGCCC TGAGGCCAGC 1200  
 AGGCGCCAG CTCAGGCAGC ACACGCCCTC ACTTAAAAAG GCCGAGGAGC GCGGGGATCC ACCTGAATCC AATTACATCT GGTGAACCTC GACATCTGAA 1300  
 ACGTTTTAAG TTACACCAAG TTCATAGCCT TTGTTAACCT TTCATGTGTT GAATGTTCAA ATAATGTTCA TTACACTTAA GAATACTGGC CTGAATTTTA 1400  
 TTAGCTTCAT TATAAATCAC TGAGCTGATA TTTACTCTTC CTTTAAAGTT TTCTAAGTAC GTCTGTAGCA TGATGGTATA GATTTTCTTG TTTCAGTGCT 1500  
 TTGGGACAGA TTTTATATTA TGTCAATTGA TCAGGTAAA ATTTTCAGTG TGTAGTTGGC AGATATTTTC AAAATTACAA TGCATTTATG GTGTCTGGG 1600  
 GCAGGGGAAC ATCAGAAAGG TTAATTTGGG CAAAATGCG TAAGTCACAA GAATTTGGAT GGTGCAGTTA ATGTTGAAGT TACAGCATTT CAGATTTTAT 1700  
 TGTGAGATAT TTAGATGTTT GTTACATTTT TAAAAATTGC TCTTAATTTT TAAACTCTCA ATACAATATA TTTTGACCTT ACCATTATTC CAGAGATTCA 1800  
 GTATTAAAA AAAAAAAT ACACGTGCGT AGTGSCATTT AAACAATATA ATATATTCTA AACACAATGA AATAGGGAAT ATAATGTATG AACTTTTTCG 1900  
 ATTGGCTTGA AGCAATATA TATATTGTAA ACAAAACACA GCTCTTACCT AATAAACATT TTATACTGTT TGTATGTATA AAATAAAGGT GCTGCTTTAG 2000  
 TTTTTTGGAA AAAAAAATA AAAAAAATA AAA 2033

FIG. 5C

10 / 108

## P1.DNA 32279

SEQ ID NO:16

Met	Pro	Gly	Ile	Lys	Arg	Ile	Leu	Thr	Val	Thr	Ile	Leu	Ala	Leu	Cys	Leu	Pro	Ser	Pro	Gly	Asn	Ala	Gln	Ala	Gln	Cys	Thr	Asn	Gly	30
				5					10					15					20				25							
Phe	Asp	Leu	Asp	Arg	Gln	Ser	Gly	Gln	Cys	Leu	Asp	Ile	Asp	Glu	Cys	Arg	Thr	Ile	Pro	Glu	Ala	Cys	Arg	Gly	Asp	Met	Met	Cys	Val	60
				35					40					45					50				55							
Asn	Gln	Asn	Gly	Gly	Tyr	Leu	Cys	Ile	Pro	Arg	Thr	Asn	Pro	Val	Tyr	Arg	Gly	Pro	Tyr	Ser	Asn	Pro	Tyr	Ser	Thr	Pro	Tyr	Ser	Gly	90
				65					70					75					80				85							
Pro	Tyr	Pro	Ala	Ala	Ala	Pro	Pro	Leu	Ser	Ala	Pro	Asn	Tyr	Pro	Thr	Ile	Ser	Arg	Pro	Leu	Ile	Cys	Arg	Phe	Gly	Tyr	Gln	Met	Asp	120
				95					100					105					110				115							
Glu	Ser	Asn	Gln	Cys	Val	Asp	Glu	Asp	Glu	Cys	Ala	Thr	Asp	Ser	His	Gln	Cys	Asn	Pro	Thr	Gln	Ile	Cys	Ile	Asn	Thr	Glu	Gly	Gly	150
				125					130					135					140				145							
Tyr	Thr	Cys	Ser	Cys	Thr	Asp	Gly	Tyr	Trp	Leu	Leu	Glu	Gly	Gln	Cys	Leu	Asp	Ile	Asp	Glu	Cys	Arg	Tyr	Gly	Tyr	Cys	Gln	Gln	Leu	180
				155					160					165					170				175							
Cys	Ala	Asn	Val	Pro	Gly	Ser	Tyr	Ser	Cys	Thr	Cys	Asn	Pro	Gly	Phe	Thr	Leu	Asn	Glu	Asp	Gly	Arg	Ser	Cys	Gln	Asp	Val	Asn	Glu	210
				185					190					195					200				205							
Cys	Ala	Thr	Glu	Asn	Pro	Cys	Val	Gln	Thr	Cys	Val	Asn	Thr	Tyr	Gly	Ser	Leu	Ile	Cys	Arg	Cys	Asp	Pro	Gly	Tyr	Glu	Leu	Glu	Glu	240
				215					220					225					230				235							
Asp	Gly	Val	His	Cys	Ser	Asp	Met	Asp	Glu	Cys	Ser	Phe	Ser	Glu	Phe	Leu	Cys	Gln	His	Glu	Cys	Val	Asn	Gln	Pro	Gly	Thr	Tyr	Phe	270
				245					250					255					260				265							
Cys	Ser	Cys	Pro	Pro	Gly	Tyr	Ile	Leu	Leu	Asp	Asp	Asn	Arg	Ser	Cys	Gln	Asp	Ile	Asn	Glu	Cys	Glu	His	Arg	Asn	His	Thr	Cys	Asn	300
				275					280					285					290				295							
Leu	Gln	Gln	Thr	Cys	Tyr	Asn	Leu	Gln	Gly	Gly	Phe	Lys	Cys	Ile	Asp	Pro	Ile	Arg	Cys	Glu	Glu	Pro	Tyr	Leu	Arg	Ile	Ser	Asp	Asn	330
				305					310					315					320				325							
Arg	Cys	Met	Cys	Pro	Ala	Glu	Asn	Pro	Gly	Cys	Arg	Asp	Gln	Pro	Phe	Thr	Ile	Leu	Tyr	Arg	Asp	Met	Asp	Val	Val	Ser	Gly	Arg	Ser	360
				335					340					345					350				355							
Val	Pro	Ala	Asp	Ile	Phe	Gln	Met	Gln	Ala	Thr	Thr	Arg	Tyr	Pro	Gly	Ala	Tyr	Tyr	Ile	Phe	Gln	Ile	Lys	Ser	Gly	Asn	Glu	Gly	Arg	390
				365					370					375					380				385							
Glu	Phe	Tyr	Met	Arg	Gln	Thr	Gly	Pro	Ile	Ser	Ala	Thr	Leu	Val	Met	Thr	Arg	Pro	Ile	Lys	Gly	Pro	Arg	Glu	Ile	Gln	Leu	Asp	Leu	420
				395					400					405					410				415							
Glu	Met	Ile	Thr	Val	Asn	Thr	Val	Ile	Asn	Phe	Arg	Gly	Ser	Ser	Val	Ile	Arg	Leu	Arg	Ile	Tyr	Val	Ser	Gln	Tyr	Pro	Phe			448
				425					430					435					440				445							

FIG. 6A

11 / 108

## P1.DNA 32292

SEQ ID NO:17

Met Arg Leu Pro Arg Arg Ala Ala Leu Gly Leu Leu Pro Leu Leu Leu Leu Leu Pro Ala Pro Glu Ala Ala Lys Lys Pro Thr Pro 30  
 1 5 25  
 Cys His Arg Cys Arg Gly Leu Val Asp Lys Phe Asn Gln Gly Met Val Asp Thr Ala Lys Lys Asn Phe Gly Gly Asn Thr Ala Trp 60  
 35 40 55  
 Glu Glu Lys Thr Leu Ser Lys Tyr Glu Ser Ser Glu Ile Arg Leu Leu Glu Ile Leu Glu Gly Leu Cys Glu Ser Ser Asp Phe Glu Cys 90  
 65 70 75 85  
 Asn Gln Met Leu Leu Glu Ala Gln Glu Glu His Leu Glu Ala Trp Trp Leu Leu Lys Ser Glu Tyr Pro Asp Leu Phe Glu Trp Phe Cys 120  
 95 100 105 115  
 Val Lys Thr Leu Lys Val Cys Cys Ser Pro Gly Thr Tyr Gly Pro Asp Cys Leu Ala Cys Gln Gly Ser Gln Arg Pro Cys Ser Gly 150  
 125 130 135 145  
 Asn Gly His Cys Ser Gly Asp Gly Ser Arg Gln Gly Asp Gly Ser Cys Arg Cys His Met Gly Tyr Gln Gly Pro Leu Cys Thr Asp Cys 180  
 155 160 165 175  
 Met Asp Gly Tyr Phe Ser Ser Leu Arg Asn Glu Thr His Ser Ile Cys Thr Ala Cys Asp Glu Ser Cys Lys Thr Cys Ser Gly Leu Thr 210  
 185 190 195 205  
 Asn Arg Asp Cys Gly Glu Cys Glu Val Gly Trp Val Leu Asp Glu Gly Ala Cys Val Asp Val Asp Glu Cys Ala Ala Glu Pro Pro 240  
 215 220 225 235  
 Cys Ser Ala Ala Gln Phe Cys Lys Asn Ala Asn Gly Ser Tyr Thr Cys Glu Glu Cys Asp Ser Ser Cys Val Gly Cys Thr Gly Glu Gly 270  
 245 250 255 265  
 Pro Gly Asn Cys Lys Glu Cys Ile Ser Gly Tyr Ala Arg Glu His Gly Gln Cys Ala Asp Val Asp Glu Cys Ser Leu Ala Glu Lys Thr 300  
 275 280 285 295  
 Cys Val Arg Lys Asn Glu Asn Cys Tyr Asn Thr Pro Gly Ser Tyr Val Cys Val Cys Pro Asp Gly Phe Glu Glu Thr Glu Asp Ala Cys 330  
 305 310 315 325  
 Val Pro Pro Ala Glu Ala Glu Ala Thr Glu Gly Glu Ser Pro Thr Gln Leu Pro Ser Arg Glu Asp Leu 353  
 335 340 345 350

FIG. 6B

12 / 108

## p1.DNA33094

SEQ ID NO:18

Met Ala Arg Ser Ala Phe Pro Ala Ala Leu Trp Leu Trp Ser Ile Leu Leu Cys Leu Leu Ala Leu Arg Ala Glu Ala Gly Pro  
 1 5 10 15 20 25 30  
 Pro Gln Glu Glu Ser Leu Tyr Leu Trp Ile Asp Ala His Gln Ala Arg Val Leu Ile Gly Phe Glu Glu Asp Ile Leu Ile Val Ser Glu  
 35 40 45 50 55 60  
 Gly Lys Met Ala Pro Phe Thr His Asp Phe Arg Lys Ala Gln Gln Arg Met Pro Ala Ile Pro Val Asn Ile His Ser Met Asn Phe Thr  
 65 70 75 80 85 90  
 Trp Gln Ala Ala Gly Gln Ala Glu Tyr Phe Tyr Glu Phe Leu Ser Leu Arg Ser Leu Asp Lys Gly Ile Met Ala Asp Pro Thr Val Asn  
 95 100 105 110 115 120  
 Val Pro Leu Leu Gly Thr Val Pro His Lys Ala Ser Val Val Gln Val Gly Phe Pro Cys Leu Gly Lys Gln Asp Gly Val Ala Ala Phe  
 125 130 135 140 145 150  
 Glu Val Asp Val Ile Val Met Asn Ser Glu Gly Asn Thr Ile Leu Gln Thr Pro Gln Asn Ala Ile Phe Phe Lys Thr Cys Gln Gln Ala  
 155 160 165 170 175 180  
 Glu Cys Pro Gly Gly Cys Arg Asn Gly Gly Phe Cys Asn Glu Arg Arg Ile Cys Glu Cys Pro Asp Gly Phe His Gly Pro His Cys Glu  
 185 190 195 200 205 210  
 Lys Ala Leu Cys Thr Pro Arg Cys Met Asn Gly Gly Leu Cys Val Thr Pro Gly Phe Cys Ile Cys Pro Pro Gly Phe Tyr Gly Val Asn  
 215 220 225 230 235 240  
 Cys Asp Lys Ala Asn Cys Ser Thr Thr Cys Phe Asn Gly Gly Thr Cys Phe Tyr Pro Gly Lys Cys Ile Cys Pro Pro Gly Leu Glu Gly  
 245 250 255 260 265 270  
 Glu Gln Cys Glu Ile Ser Lys Cys Pro Gln Pro Cys Arg Asn Gly Gly Lys Cys Ile Gly Lys Ser Lys Cys Lys Cys Ser Lys Gly Tyr  
 275 280 285 290 295 300  
 Gln Gly Asp Leu Cys Ser Lys Pro Val Cys Glu Pro Gly Cys Gly Ala His Gly Thr Cys His Glu Pro Asn Lys Cys Gln Cys Gln Glu  
 305 310 315 320 325 330  
 Gly Trp His Gly Arg His Cys Asn Lys Arg Tyr Glu Ala Ser Leu Ile His Ala Leu Arg Pro Ala Gly Ala Gln Leu Arg Gln His Thr  
 335 340 345 350 355 360  
 Pro Ser Leu Lys Lys Ala Glu Glu Arg Asp Pro Pro Glu Ser Asn Tyr Ile Trp  
 365 370 375 379

FIG. 6C

SEQ ID NO 22	2305118	1	GCCGCTTTGGATACCAGATGGATGAACCAACCAATGTGTGGATGTGGAC	GAGTGTGCAACAGATTCACCAAGTCCACCAAGTGCACCCACCCAGATCTGCATCAA
SEQ ID NO 23	2544914	1	ACCAATGTGTGGATGTGGAC	GAGTGTGCAACAGATTCACCAAGTCCACCAAGTGCACCCACCCAGATCTGCATCAA
SEQ ID NO 19	<DNA28726>	1	GCCGCTTTGGATACCAGATGGATGAACCAACCAATGTGTGGATGTGGAC	GAGTGTGCAACAGATTCACCAAGTCCACCAAGTGCACCCACCCAGATCTGCATCAA
SEQ ID NO 24	2305118	101	TACTGAAGGGGGGTACACCTGCTCTGCACCGACGGATATTGGCTTCTGG	AAGGCCAGTGTCTAGACATTTGATGAATGTCGCTATGGTTACTGCCAGCAG
SEQ ID NO 25	2544914	71	TACTGAAGGGGGGTACACCTGCTCTGCACCGACGGATATTGGCTTCTGG	AAGGCCAGTGTCTAGACATTTGATGAATGTCGCTATGGTTACTGCCAGCAG
SEQ ID NO 26	1682522	1	CTCGAGCCGCGGATATTGGCTTCTGG	AAGGCCAGTGTCTAGACATTTGATGAATGTCGCTATGGTTACTGCCAGCAG
SEQ ID NO 27	424333	1	CACCGACGGATATTGGNTTCTGG	AAGGCCAGTGTCTAGACATTTGATGAATGTCGCTATGGTTACTGCCAGCAG
SEQ ID NO 28	640534	1	GCTTCTGG	AAGNCCAGTGTCTAGACATTTGATGAATGTCGCTATGGTTACTGCCAGCAG
SEQ ID NO 29	2211568	1	TACTGAAGGGGGGTACACCTGCTCTGCACCGACGGATATTGGCTTCTGG	AAGGCCAGTGTCTAGACATTTGATGAATGTCGCTATGGTTACTGCCAGCAG
SEQ ID NO 30	<DNA28726>	101	TACTGAAGGGGGGTACACCTGCTCTGCACCGACGGATATTGGCTTCTGG	AAGGCCAGTGTCTAGACATTTGATGAATGTCGCTATGGTTACTGCCAGCAG
SEQ ID NO 31	2305118	201	CTCTGTGCGAATGTTCTCTGGATCCTATTCTTGTACATGCAACCCCTGGTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 32	2544914	171	CTCTGTGCGAATGTTCTCTGGATCCTATTCTTGTACATGCAACCCCTGGTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 33	1682522	77	CTCTGTGCGAATGTTCTCTGGATCCTATTCTTGTACATGCAACCCCTGGTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 34	424333	74	CTCTGTGCGAATGTTCTCTGGATCCTATTCTTGTACATGCAACCCCTGGTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 35	640534	59	CTCTGTGCGAATGTTCTCTGGATCCTATTCTTGTACATGCAACCCCTGGTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 36	2211568	8	CTCTGTGCGAATGTTCTCTGGATCCTATTCTTGTACATGCAACCCCTGGTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 37	1436024	1	CCTGGTTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 38	<DNA28726>	201	CTCTGTGCGAATGTTCTCTGGATCCTATTCTTGTACATGCAACCCCTGGTT	TACCCCTCAATGAGGATGGAGGTCTTG
SEQ ID NO 39	1682522	177	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTT	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 40	424333	174	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTT	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 41	640534	159	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTT	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 42	2211568	108	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTT	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 43	1436024	59	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTT	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 44	W24885	1	ACACCTACGGCTCTTTCATC	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 45	1600521	1	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTTTCATC	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 46	732577	1	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTTTCATC	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG
SEQ ID NO 47	<DNA28726>	301	CCGAGAACCCCTGCGTGCAACCTGCGTCAACACCTACGGCTCTTTCATC	TGCGGCTGTGACCCAGGATATGAATTTGAGGAAGATGGCGTTCAATTGCAG

FIG. 7A-1

**FIG. 7A-2**

SEQ ID NO 37	2398238	1	CGGACAGGTGTGACTCCAGCTGTGTGGGCTGCACAGGGGAAGGCCCCAGGA	AACTGTAAAGAGTGTATCTCTGGCTACGGAGGAGACGACAGTGTGC
SEQ ID NO 38	1842628	1	ATTTCATAATCCG	CTGGCGCTCTGCTGAGCAAGTAATTAATAACGGAGCTCTTTTCATTTT
SEQ ID NO 39	2191592	1	GGAGGCCCCAGGA	AACTGTAAAGAGTGTATCTCTGGCTACGGAGGAGACGACAGTGTGC
SEQ ID NO 40	1932631	1	CCCAGGA	AACTGTAAAGAGTGTATCTCTGGCTACGGAGGAGACGACAGTGTGC
SEQ ID NO 41	AA195267	1		ATTCCGACAGGAGGACAGTGTGC
SEQ ID NO 42	H99879	1		GCGAGGAGACGACAGTGTGC
SEQ ID NO 43	AA195084	1		GGACAGTGTGC
SEQ ID NO 21	<DNA28730>	1	CGGACAGGTGTGACTCCAGCTGTGTGGGCTGCACAGGGGAAGGCCCCAGGA	AACTGTAAAGAGTGTATCTCTGGCTACGGAGGAGACGACAGTGTGC
		101	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	2398238	64	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	1842628	64	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	2191592	64	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	1932631	58	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	AA195267	25	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	H99879	24	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	AA195084	12	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
	1700782	1	CGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
SEQ ID NO 44	<DNA28730>	101	AGATGTGACGAGTGTCTCACTAGCAGAAAAAACCTGTGTGAGGAAAAACG	AAAACCTGCTACATACTCCAGGAGCTACGTCCTGTGTGTCTCTGACGGC
		201	TTCGAAGAAACGGAA	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCTGTAAAT
	2398238	164	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GAGGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	1842628	164	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GAGGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	2191592	164	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GAGGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	1932631	158	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GAGGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	AA195267	125	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GGAGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	H99879	124	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GGAGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	AA195084	112	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGN-A-GAGGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	1700782	92	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GAGGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
	<DNA28730>	201	TTCGAAGAAACGGAAAGATGCCCTGTGTG-CGCGCGGCA-GAGGCTGAAGCC	ACAGAAGGAGAAAGCCCGACACACAGCTGCCCTCCCGTGAAAGACCT
		225	TGCCGGACTTACCCCTTTAAATTAATTCAGAAGGATGCCGTGGAATAATGT	GGCCCTGAGGATGCCGTCTCCTGAGTGACAGCGCGGGGAGAGGCTGC
	AA195267	224	TGCCGGACTTACCCCTTTAAATTAATTCAGAAGGATGCCGTGGAATAATGT	GGCCCTGAGGATGCCGTCTCCTGAGTGACAGCGCGGGGAGAGGCTGC
	H99879	212	TGCCGGACTTACCCCTTTAAATTAATTCAGAAGGATGCCGTGGAATAATGT	GGCCCTGAGGATGCCGTCTCCTGAGTGACAGCGCGGGGAGAGGCTGC
	AA195084	192	TGCCGGACTTACCCCTTTAAATTAATTCAGAAGGATGCCGTGGAATAATGT	GGCCCTGAGGATGCCGTCTCCTGAGTGACAGCGCGGGGAGAGGCTGC
	1700782	301	TGCCGGACTTACCCCTTTAAATTAATTCAGAAGGATGCCGTGGAATAATGT	GGCCCTGAGGATGCCGTCTCCTGAGTGACAGCGCGGGGAGAGGCTGC
	<DNA28730>			
	AA195267	325	CTGCTCTTAACGGTTGATTTCTCATTTGTCCCTTAAACAGCTGCAATTTCT	TGGTTGTTCTTAAACAGACTTGATATATTTTGTATACAGTTCTTTTGTAAATAA
	H99879	324	CTGCTCTTAACGGTTGATTTCTCATTTGTCCCTTAAACAGCTGCAATTTCT	TGGTTGTTCTTAAACAGACTTGATATATTTTGTATACAGTTCTTTTGTAAATAA
	AA195084	312	CTGCTCTTAACGGTTGATTTCTCATTTGTCCCTTAAACAGCTGCAATTTCT	TGGTTGTTCTTAAACAGACTTGATATATTTTGTATACAGTTCTTTTGTAAATAA
	<DNA28730>	401	CTGCTCTTAACGGTTGATTTCTCATTTGTCCCTTAAACAGCTGCAATTTCT	TGGTTGTTCTTAAACAGACTTGATATATTTTGTATACAGTTCTTTTGTAAATAA
		425	AATTGACCATTTAGGTAAA	
	AA195267	424	AATTGACCATTTAGGTAAA	
	H99879	412	AATTGACCATTTAGGTAAA	
	AA195084	501	AATTGACCATTTAGGTAAA	
	<DNA28730>			

FIG. 7B



SEQ ID NO 45	298	CNAANNANCTGGNNAANNNNNNTTGANCAANNANCTGGNNA	ANCCNNNTTGAAGAAGTTCAACCT-GGTAAAGNCCCAANCTGAATTCNC
SEQ ID NO 46	97	TCCCCNNNACCCCAANNCAACTTTTAAAAAATTNNCCCAACCAANTTG	ATTCCAGGGGGTCCCGAATGG-GNGGCCCTTTTNTTATTGAAAAAAN
SEQ ID NO 47	53	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	GTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 48	51	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	NGTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 49	47	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	GGTTTTNNNAAGGNTTCCGG-TGTNCCNTGGNTCCANGNCCCTC
SEQ ID NO 50	44	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	TGTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 51	1	TTTCGAATGCCCACTGAGTCCAGCGGGTCCGAATGGAGGCTTT	TGTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 52	1	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	TGTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 53	1	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	TGTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 54	1	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	TGTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 55	1	AAACATGTCAACAAGCTGAGTCCAGCGGGTCCGAATGGAGGCTTT	TGTAATGAAGAAGCATCTGCGAG-TGTCTGATGGTTCACGGACCTC
SEQ ID NO 20	<DNA28760>		
SEQ ID NO 21	398	AAGTTGAGGGTCCCGTGAACCCNCCGATGATGAATGGTGGACNTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 22	197	NAACCTCGGATTNTCTTGATGGTTCNANGGACCTCCCTGTGAGAAA	CCCCNTGTACCCCCCGATGATGAGTGGTGGCCTTTGTGTGANCACCTG
SEQ ID NO 23	153	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 24	151	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 25	147	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 26	144	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 27	101	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 28	92	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 29	39	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 30	39	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 31	1	CTTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 32	101	ACTGTGAGAAAGCCCTTGTACCCACGATGATGAATGGTGGACCTTGT	GTGACTCCTGTTTCTGCATCTGCCACCTGGATTTCTATGGAGTGAACCTG
SEQ ID NO 33	498	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTNCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 34	297	GGTNAACATCTGCCACCTGGATTTCTATGAGTNGCTGTGNCAAN	CACACTCNTCANCCACCTCTTAAATGGAGGACCTGTTNCTACCTCGA
SEQ ID NO 35	253	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTNCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 36	251	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 37	247	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 38	244	AGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 39	201	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 40	192	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 41	139	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 42	139	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 43	87	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 44	1	ATNTGGAACGAGGAGTGAACCTGACCAAG	CAAACCTGCTCAACCACTGCTTTAATGAGGACCTGTTTCTACCTGGA
SEQ ID NO 45	1	GGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 46	1	GGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 47	1	GGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA
SEQ ID NO 48	201	TGACAAAGCAAACTGCTCAACCACTGCTTTAATGAGGACCTGTTTCT	ACCCTGGAAAATGATTTTCCCTCCAGGACTAGAGGGAGAGCAGTGTGAA

FIG. 7C-1

17 / 108

W27896 598 ATCAGCAAAATGCCCAACACCCCTGTCGAAATGGAGG-TAAATGCATTGG-T AAAAGCAAAATGTAA-GT-GTTCC-AAAGG-TTACC-AGGG-AGACCTCT-  
 W27851 397 AAATGTANATTTNCCCAGGACTAGAGGGAGAGNA-GTGTGNNATCAC-C AAATATCCCACCN-CT-CTGCG-GAAAT-TGGG-AGG-CAATGGTA-  
 W22553 353 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG-TAAATGCATTGG-T AAAAGCAA-TGTAA-GT-GTTCC-AAAGG-TTACC-AGGG-AGACNCTT-  
 W23268 351 ATCAGCAAAATGCC-ACAACCTGTGCAATGGAGG-TAAATGCATTGG-T AAAAGNAA-TGTAA-GT-GT-CC-AAAGG-TAAC--AGGG-AGACCTNT-  
 W28670 344 TTCACNANCCCCCAACACCCGNGNNAATGGANG-TATNTACATCNN-T AATAGTATTTCCNCC-GT-GTACC-AAAGG-TGACC-ACTG-AGNACNCT-  
 W27944 301 ATCAGNAAATGCCCAACACCCCTGTGCAATGGAGG-TAAATGCATTGG-T AAAAGCAATGTAA-GTTGTTCCCAAGGGTTACCCAGGG-AGACCTCT-  
 R55894 292 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG-TAAATGCATTGG-T AAAAGCAAAATGTAAAGT-GTTCC-AAAGG-TTACC-AGGGAGACCTCT-  
 660500 239 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG-TAAATGCAT  
 662092 239 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG  
 1682022 187 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG-TAA  
 W37154 82 AAATGTATTTGCCCTCCAGGACTAGAGGAGGACA-GTGTGAAATCAG-C AAATGCCCAACAC-CT-GTCCA-AAATGG-AGTA-AATG-CAATGGTA-  
 1577139 67 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG-TAAATGCATTGG-T AAAAGCAAAATGTAA-GT-GTTCC-AAAGG-TTACC-AGGG-AGACCTCT-  
 W38638 56 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG-TAAATGCATTGG-T AAAAGCAAAATGTAA-GT-GTTCC-AAAGG-TTACC-AGGG-AGACCTCT-  
 301 ATCAGCAAAATGCCCAACACCCCTGTGCAATGGAGG-TAAATGCATTGG-T AAAAGCAAAATGTAA-GT-GTTCC-AAAGG-TTACC-AGGG-AGACCTCT-  
 <DNA28760>  
 W27896 698 GTT-CAAAAGCCTGTCT-GCG-AGCCTGG-CT-GTGG-T-GCACA-TGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-C  
 W27851 497 AAA-GCAAAATCTGTCT-GCG-AGCCTGG-CT-GTGG-T-GCACA-TGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-C  
 W22553 453 NTC-CAAAAGCCTGTCT-GCG-AGCCTGG-CT-GTGG-T-GCACA-TGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-C  
 W23268 451 --T-CAAAAGC-TGTNT-GGA-GCTGTGG-GG-GGAA-T-GACNG-CAT-A ACC-AAATNCA-ANTAGAGGTG-NATGAG-NAT-GATAAGTC-GAGC-N  
 W28670 444 NNT-CATAGNCTNCT-GTGGAGGCTGG-AT-ATGG-T-GCACA-TCG-T ANC-CCCCACA-AATCCCAACAAATANCAA-NGT-CAAAATGG-GTGN-N  
 W27944 401 GTT-CAAAAGCCTGTCTGTGG-AGCCTGG-TTGTGGT-GCACAATGG-A ACCCTGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-G  
 W25894 392 GTTCAAAAGCCTGTCTGTGG-AGCCTGG-TTGTGGT-GCACAATGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-G  
 W37154 182 AAA-GCAAAATCTGTCT-GCG-AGCCTGG-CT-GTGG-T-GCACA-TGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-C  
 1577139 167 GTT-CAAAAGCCTGTCT-GCG-AGCCTGG-CT-GTGG-T-GCACA-TGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-C  
 W38638 156 GTT-CAAAAGCCTGTCT-GCG-AGCCTGG-CT-GTGG-T-GCACA-TGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-C  
 401 GTT-CAAAAGCCTGTCT-GCG-AGCCTGG-CT-GTGG-T-GCACA-TGG-A ACC-TGCCATG-AACCCCAACAAA-TGCCAA-TGT-CAAGAAGG-TTGG-C  
 W27896 798 A-TGGAAGACACTGCAATATAAAGGGACGAGCCAGCCTCATACATGCCCT GA  
 W27851 597 A-TGGAAGACACTGCAATATAAAGGGACGAGCCAGCCTCATACATGCCCT GAGTAAGNAA  
 W22553 553 A-TGGAAGNANTTGAATAAGGCTCGAGGCAACCCAAATTTCTTNGCC ANAGGCCAGTTAAGAAANAAANCCTCCATTAAANGGCGGGGGGGTCCC  
 W23268 551 C-CNTNATNCTAGCAAGGCGCGTAGAGNANCNTAAAGGGGGGGGGCCT NTCNNTNGTTCGGTNNNNNNNNNGNACCTNTTGCANNNTCTTNC  
 W28670 544 N-AATAAAGGAACTGNAAGGANGGNNCCNCCNCAATCCCAATGG CTCACNCGTTCACACCGCCTTACCAACTAAGGGGGGCAANNNGNTT  
 W27944 501 AATGGNGAACTNNAATAAANGGTACGGAGGCAAGCCCAANCAATGCC NNTGNTTCAATTGGTTTCAAGTGTNTTANAAAAANNNNNNTN  
 R55894 492 N-TGGAAGACACTGCAATATAAAGGGACGAGCCAGCCTCATACATGCCCT  
 W37154 282 A-TGGAAGACACTGCAATATAAAGGTACGAAGCCAGCCTCATACATGCCCT TAGGCCAAGCAGCGGCCAGCTCAGGCAGCACACGCCCTTCACTTAAAAAG  
 W38638 256 A-TGGAAGACACTGCAATATAAAGGTACGAAGCCAGCCTCATACATGCCCT GAGNCCAAGCAGN-GCCAGCTCAGGCAGCACACGCCCTTCACTTAAAAAG  
 501 A-TGGAAGACACTGCAATATAAAGGTACGAAGCCAGCCTCATACATGCCCT TAGGCCAAGCAGCGGCCAGCTCAGGCAGCACACGCCCTTCACTTAAAAAG  
 <DNA28760>  
 W22553 653 CTTCTTTGGTCCNAAAAA-AAA-AA-AAAAAANNTTNNNNNNNNN NNNNNNNNNNNNNCNCNTCNC  
 W28670 644 TACCTNNNNNTTNNNNNNNN-NN-TCCCGGTTNNNTTNNAAANN NTTNCCAGATTNNCTCNAANNNNTTNCAGNNTTNCCTNN  
 W27944 601 CCCNGTNNNTTNNAAAAANNN-NN-TTNCNNNNNTTNNCAAAAAAN NNN  
 W37154 382 GCGAGGAGCGGGGGATCCACC-TG-AATCCATTTACCAAGCTTGA ACTCCGACATCTGGAAACCGTTTAAAGTACACCAAGTTCATAG  
 W38638 356 GCGAGGAGCGGGGGATTCACNTGGAATCCAAAT-ACATCTGG-TGA ACTCCGACATCTG-AAAAGTAAAGTTACACCAAGTTCATAG  
 601 GCGAGGAGCGGGGGATCCACC-TG-AATCCAAATTTACATCTGGGTGA ACTCCGACATCTGGAACCGTTTAAAGTACACCAAGTTCATAG  
 <DNA28760>

FIG. 7C-2

18 / 108

**28726.p**

SEQ ID NO 60

GGGTACACCTGCTCCTGCACCGACGGATATTGGCTTCTGGAAGGCC

**FIG. 8A**

**28726.f**

SEQ ID NO 61

ACAGATTCCCACCAAGTGCAACC

**FIG. 8B**

**28726.r**

SEQ ID NO 62

CACACTCGTTCACATCTTGGC

**FIG. 8C**

19 / 108

**28730.p**  
SEQ ID NO 63

AGGGAGCACGGACAGTGTGCAGATGTGGACGAGTGCTCACTAGCA

**FIG. 9A**

**28730.f**  
SEQ ID NO 64

AGAGTGTATCTCTGGCTACGC

**FIG. 9B**

**28730.r**  
SEQ ID NO 65

TAAGTCCGGCACATTACAGGTC

**FIG. 9C**

20 / 108

**28760.p**

SEQ ID NO 66

CCCACGATGTATGAATGGTGGACTTTGTGTGACTCCTGGTTTCTGCATC

**FIG. 10A**

**28760.f**

SEQ ID NO 67

AAAGACGCATCTGCGAGTGTCC

**FIG. 10B**

**28760.r**

SEQ ID NO 68

TGCTGATTTCACACTGCTCTCCC

**FIG. 10C**

Frame Score Match Pct  
+3 962 170 48

GEN12205 Epidermal growth factor-like protein S1-5 -

GEN12205 Epidermal growth factor-like protein S1-5 - human (497 aa)

Score = 962 (338.6 bits), Expect = 4.3e-96, P = 4.3e-96

Identities = 170/353 (48%), Positives = 225/353 (63%), at 735,148, Frame = +3

DNA32279	735	APPLSAPNYPTISRPLICRFGYQMDSESNQVDVDECATDSHQCNPTQICINTEGGYTCSC
		* *
GEN12205	148	ADPQIPSNP--SHRIQCAAGYEQSEHNVCQDIDECTAGTHNCRADQVCINLRGSFACQC
SEQ ID NO:69		
DNA32279	915	TDGYWLLEGQCLDIDECRYG-YCQQLCANVPGSYSCNPGFTLNEDGRSCQDVNECATE
		** ** ** ** ** ** ** ** ** ** ** ** *
GEN12205	206	PPGYQKRGEQCVDIDECTIPPYCHQRCVNTPGSFYQCSPGFQLAANNYTCVDINECDAS
DNA32279	1092	NPCVQTCVNTYGSILICRCDPGYELEDGVHCSDMDECSFSEFLCQHECVNQPGTYFCSCP
		* *
GEN12205	266	NQCAQQCYNILGSFICQCNQGYELSSDRLNCEIDECRTSSYLQYQCVNEPGKFSMCP
DNA32279	1272	PGYILLDDNRSCQDINECEHRNHTCNLQQTCTYNLQGGFKCIDPICEEPYLRISDNRCMC
		** .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * ..
GEN12205	326	QGYQVVR-SRTCQDINECETTNE-CREDEMCMWYHGGFRCYPRNPCQDPYILTPENRCVC
DNA32279	1452	PAENPGCRDQPFITLYRDMDVSGRSVPADIFQMQAATTPYCGAYYIFQIKSGNEGREFYM
		* *
GEN12205	384	PVSNAMELPQSIYVKYMSIRSDRSVPSDIFQIQATTIVANTINTFRIKSGNENGEFYL
DNA32279	1632	RQTGPISATLVMTTRPIKGPRIQLDLEMITVNTVINFRGSSVIRLRIYVSQYPPF
		*** * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * .. * ..
GEN12205	444	RQTSPVSAMLVLVKSLSGPREHIVDLEMLTVSSIGTFRFTSSVLRILTIVGPFPSF

FIG. 11-1A

**FIG. 11-1B**

```

DNA32279      426 ISSSRVLDMPGIKRLTVTILAL--CLPSPGNAQAQCTNGFDLDRQSGQCLDIDECRTIP
      . ** * * . * * * * . * * . * * * * * * *
GEN12205      289 LSSDR-LNCEDIDECRTSSYLCOYQCVNEPGKFCMCPQGYQVVR-SRTCQDINECETTN

DNA32279      600 EACRGDMMCVNQNGGYLCIPRTNPVYRGPY
      * * * * * * * * * * * * * *
GEN12205      347 E-CREDEMCMWNYHGGFRCYPR-NPC-QDPY

```

**FIG. 11-2**



PAC6	RAT	Serine protease pc6 precursor - rattus norv...	Frame	Score	Match	Pct
			+2	209	67	36

**PAC6 RAT Serine protease pc6 precursor - rattus norvegicus (915 aa)**

**Score = 209 (73.6 bits), Expect = 2.3e-12, Sum P(2) = 2.3e-12**

**Identities = 67/186 (36%), Positives = 87/186 (46%), at 473,722, Frame = +2**

DNA32292 473 CLA-CQGGSRPCSGN--GHCSGDGSR-QGDGSC-RCHMGY--QGPLCT-DCMDGYFSSL

**PAC6 RAT** 722 CVAQCEGSYQDIKKNICGKCSENCKTCTGFHNCTECKGGLSLQGSRCSVTCEDGQFFS-

**SEQ ID NO:70**

DNA32292 629 RNETHSICTACDESKCTCSGLTNRDCECEVGWVLDEGACVDVDECAAEPPPCSAAQFCK

**PAC6 RAT** 781 ---GHD-CQPCRFCATCAGAGADGCINCTEGYVMEEGRCVQ-----SCSVSYLD

DNA32292 809 NA-NGSY-TCEECDSSVCVTGEGPGNCCKECISGYAREHGQCA-----DVDECSLAECT

PAC6 RAT 828 HSLEGGYKSKRCRCDNSCLTCNCGPFKNCCSPSGYLLDLGMCQMGAIKCKDATEESWAEGG

DNA32292 965 -C--VRKNENCYNTPGSYVC

\*  
\*  
\*  
\*  
\*  
\*

**PAC6 RAT 888 FCMLVKNNLCQRKVLQQLC**

**FIG. 12A-1A**

25 / 108

Score = 197 (69.3 bits), Expect = 5.0e-11, Sum P(2) = 5.0e-11  
 Identities = 62/199 (31%), Positives = 85/199 (42%), at 437,670, Frame = +2

```

DNA32292  437 KVC---CSPGTYGPDCLACQGGSQRPCSGNGHCSGDGSRQGDGSCRCHMGY-----QGPLC
PAC6_RAT      .. * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
        670 RIVSSCPPGHEHADKKRC-----RKAPN--CESCFGSHADQCLSKYGYFLNEETSSC

DNA32292  596 T-DCMDGYFSSLRNETHSICTACDESKTCSGLTNRDCGECEVGWVLDEGACVDVDECAA
PAC6_RAT      * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
        723 VAQCPEGSYQDIKK---NICGKCSENCKTCTGFHN--CTECKGGLSL-QGS-----RCSV

DNA32292  773 EPPPCSAAQFCKNANGSYTCEECDSSCVGCTGEGPGNCKECISGYAREHGGQCADVDECSL
PAC6_RAT      * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
        772 T---CEDGQFFSG-----HDCQPCHRFCATCAGAGADGCINCTEGYVMEEGRCVCQSCSVSY

DNA32292  953 -----AEKTCVRKNENCY--NTPGSYVCV-CPDGF
PAC6_RAT      * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
        825 YLDHSLEGGYKSKRCRCDNSCLTCNGPGFKNCSSCPSGY
  
```

FIG. 12A-1B

26 / 108

Score = 185 (65.1 bits), Expect = 1.1e-09, Sum P(2) = 1.1e-09  
 Identities = 70/216 (32%), Positives = 90/216 (41%), at 443,647, Frame = +2

```

DNA32292  443 CCSPGT-YGPDCLACQ---GGSQRPCSGN---GHCSGDGSR--QGDGSCRCHMGYQGPLC
          *  ** . ***          *  * . ** *  *  *  *  *  *  *  *
PAC6_RAT  647 CDGPGPDHCTDCLHYHYKLNNTTRICVSSCPPGHFADKKRCRKCAPNCESCFGSHADQC

DNA32292  596 TDCMDGYFSSLRNETHS-----ICTACDESKTCSGLTNRDCGECEVGW
          *  *** *  ** *          ** *  *  *  *  *  *  *  *  *
PAC6_RAT  707 LSCKYGYF--LNEETSSCVAQCPEGSYQDIKKNICGKCSENCKTCTGFHN--CTECKGGL

DNA32292  728 VLDEGACVDVDECAAEPPPCSAQAQFCKNANGSYTCEECDSSCVGCTGEGPGNCKECISGY
          *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
PAC6_RAT  763 SL-QGS-----RCSVT---CEDGQFFSG----HDCQPCHRFCAATCAGAGADGCINCTEGY

DNA32292  908 AREHGQCADVDECSLAE-----KTCVRKNENCY--NTPGSYVCV-CPDGF
          *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
PAC6_RAT  810 VMEEGRC--VQSCSVSYIYLDHSLEGGYKSKCRCDNSCLTCNGPGFKNCSSCPSGY
  
```

FIG. 12A-2A

Score = 93 (32.7 bits), Expect = 8.8, Sum P(2) = 1.0  
Identities = 37/132 (28%), Positives = 49/132 (37%), at 659,638, Frame = +2

DNA32292	659	CDESKT--CSGLTNRDCGEC-EVGWVLDEGACVDVDECAAEPPPCSAQFCKNANGSYT
PAC6_RAT	638	CDPECSEVGCDGPGPDHCTDCLHYHYKLKNNTRICVSSC---PP-----GHF--HADKK-R
DNA32292	830	CEECDSSCVGCTGEGPGNCKECISGYA--REHGQC-ADVDECS---LAECTCVRKNENCY
PAC6_RAT	688	CRKCAPNCESCFGSHADQCLSCKYGYFLNEETSSSCVAQCPEGSYQDIKKNICGKCSENC
DNA32292	992	NTPGSYVCV-CPDGFEEEDAC
PAC6_RAT	748	TCTGFHNCTECKGGLSLQGSRC

Score = 42 (14.8 bits), Expect = 2.3e-12, Sum P(2) = 2.3e-12  
Identities = 10/15 (66%), Positives = 10/15 (66%), at 90,5, Frame = +3

DNA32292	90	WGSCRFCCCCRRRR
PAC6_RAT	5	WGS-R---CCRPGR

FIG. 12A-2B

	Frame	Score	Match	Pct
FBLC_MOUSE Fibulin-1, isoform c precursor - mus musc...	+2	206	65	31

FBLC\_MOUSE Fibulin-1, isoform c precursor - mus musculus (685 aa)

Score = 206 (72.5 bits), Expect = 1.2e-12, P = 1.2e-12

Identities = 65/213 (30%), Positives = 89/213 (41%), at 449,211, Frame = +2

```

DNA32292      449 SPGTYGPDCLACQGSQRPCSGNGHCSGDSRQDGSRCRCHMGYQGPLCTDCMDGYFSSL
          * * * * *
FBLC_MOUSE    211 SDGVSCEDINECITGSHNCRLGESCINTVGSFRCQDSSCGTGVELTEDNNCKD--IDEC
          * * * * *
SEQ ID NO:71
    
```

```

DNA32292      629 RNETHSICTACDESK-TCSGLTNRDCGECEVGWLDE-GACVDVDECAAEPPPCSAAQF
          * * * * *
FBLC_MOUSE    269 ETGIHN-CPP-DFICQNTLGSFRCRPRKLQCKSGFIQDALGNCIDINECLISAPCPVGQT
          * * * * *
    
```

```

DNA32292      803 CKNANGSYTCEECDSSCVGCTGEGPGNCKECISGY-AREHG-QCADVDECSLAECTCVRK
          * * * * *
FBLC_MOUSE    327 CINTEGSYTCQK-----NVPNCGRGYHLNEEGTRCVDVDECAPPAEPC-GK
          * * * * *
    
```

```

DNA32292      977 NENCYNTPGSYVCVCPDGF--EETEDACVPPAEAEATEG
          * * * * *
FBLC_MOUSE    372 GHCLNSPGSFRCECKAGFYFDGIGISRTCVDINECQRYPG
          * * * * *
    
```

FIG. 12B-1

Score = 98 (34.5 bits), Expect = 0.83, P = 0.56  
Identities = 38/128 (29%), Positives = 56/128 (43%), at 701,310, Frame = +2

DNA32292	701	DCGEC-EVGWVLD	EG-ACVDVD	---ECAAEP	PPCSAAQ	FCKNANG	SYTC	---EECDSSCV
		* **	*	* **	*	*	*	* **
FBLC_MOUSE	310	DINECLSISAP	CPVGQTC	INTEGSYTC	QKNVPNC	GRG-YHLNE	EGT-RCVDV	DECAPPAE
		*	*	*	*	*	*	*
DNA32292	857	GCTGEG	-----PGNCK-ECISGYARE	--HGQCAD	VDECS-LAEKTC	VRKNENC	YNTPGS	
		* **	**	**	**	*	*	*
FBLC_MOUSE	368	PC-GKGHHCL	NSPGSFRCECKAGFY	FDGISRT	CVDINECQRY	PGR	LCGHK	---CENTPGS
		*	*	*	*	*	*	*
DNA32292	1007	YVCVCPD	GFEE	TE	D			
		*	*	*	*	*	*	*
FBLC_MOUSE	424	FHCSCS	AGFR	LSVD				
		*	*	*	*	*	*	*

FIG. 12B-2

A43902	tenascin - eastern newt (fragment).. +2	Frame	Score	Match	Pct
			336	59	36

Score = 336 (118.3 bits), Expect = 1.2e-26, P = 1.2e-26  
Identities = 59/163 (36%), Positives = 79/163 (48%), at 674,65, Frame = +2

DNA33094	674	CQQAECPPGGCRNGGFCNERRICECPDGFHGHCEKALCTPRCMNGGLCVTPGFCICPPGF
A43902	65	CSELICPNDCFDRGRCING-VCFCDEGFTGEDCGELTCPNNCNRGRVCVN-GLCVCDDGF
SEQ ID NO:72		
DNA33094	854	YGVNCDKANCSTTCFNGGTCFYPGKICICPPGLEGEQCEISKCPQPCRNGGKIGKSKKC
A43902	123	QGDDCSELRCPNDCNDRGRVCVN-GKCVCKEGFMGEDCADLRCPNDCNNRGRVCVN-GQCVC
DNA33094	1034	SKGYQGDLCSPVCEPGCGAHGTCHPEPNKCQCQEGWHGRHCNK
A43902	181	DEGFMGEDCSDLRCPGDCNNRGRVCVN-GQCVCDEGFRGEDCGE

FIG. 13A-1

**Score = 135 (47.5 bits), Expect = 0.00013, P = 0.00013**  
**Identities = 24/67 (35%), Positives = 35/67 (52%), at 962,3, Frame = +2**

DNA 33094	674	CQQAECPPGGCRNGGFCNERRICECPDGFHGPHEKALCTPRCMNGGLCVTPGFICPPGF
		* ** * * * . * * * * . * * * * . * * * * . * * * * . * * * * .
A43902	282	CSELRCPDDCNDRCRCINGQ-CVCAEGFTGENCDSLACLNNCNDRGLCVN-GQCVCEEGF
		. * * * * . * * * * . * * * * . * * * * . * * * * . * * * * .
DNA 33094	854	YGVNCDKAN
		* . * . .
A43902	340	LGEDCSEVS

**Score = 135 (47.5 bits), Expect = 0.00013, P = 0.00013**  
**Identities = 24/67 (35%), Positives = 35/67 (52%), at 962,3, Frame = +2**

DNA33094	962	CEISKCPQPCRNGGKCIGKSKCKCSKGYQGDLCSPVCEPCCGAHGTCHEPNKCQCQEGW
		* * * * * . * * * * . * * * * . * * * * . * * * * . * * * * .
A43902	3	CGQEICQVECFGKCVN-GQCVCEGFTGEDCSEPRCPNNCNDRGRCVE-DECVCDEGE
		. * * * * . * * * * . * * * * . * * * * . * * * * . * * * * .
DNA33094	1142	HGRHCNK
		* * . .
A43902	61	TGDDCSE

**FIG. 13A-2**



Frame    Score    Match    Pct  
+2       331       61       37

HSTNX12\_1    tenascin-X precursor - Homo sapiens

Score = 331 (116.5 bits), Expect = 6.7e-26, P = 6.7e-26  
Identities = 61/164 (37%), Positives = 74/164 (45%), at 674,247, Frame = +2

DNA33094	674	CQQAECPPGGCRNGGFCNERRICEPDGFHGHCEKALCTPRCMNGGLCVTPGFCICPPGF
		* * * * *
HSTNX12_1	247	CSQSRSCPRGCSQGRCEGGR-CVCDPGTYGDDCGMRSCPRGCSQGRGCEN-GRCVCNPGY
		* * * * *
SEQ ID NO:73		
DNA33094	854	YGVNCDKANCSTTCFNGGTCFYPGKICPPGLEGEQCEISKCPQPCRNGGKCIKSKKCC
		* * * * *
HSTNX12_1	305	TGEDCGVRSCPRGCSQGRGC-KDGRVCVDPGYTGEDCGTRSCPWDCGEGGRCVD-GRCVC
		* * * * *
DNA33094	1034	SKGYQGDLCSPVCEPGCGAHGTCHEPNKCQCQEGWHGRHCNKR
		* * * * *
HSTNX12_1	363	WPGYTGEDCSTRTCPRDCRGRGC-EDGEICIDTGYSGDDDCGVR
		* * * * *

FIG. 13B-1A

33 / 108

Score = 324 (114.1 bits), Expect = 3.9e-25, P = 3.9e-25  
 Identities = 63/171 (36%), Positives = 74/171 (43%), at 674,464, Frame = +2

DNA 33094	674	CQQAECPGGCRNGGFCNERRICECPDGFHGHCEKALCTPRCMNGGLCVTPGFICPPGF
		* *** ** * * * * * * * * * * * * * * * * * *
HSTNX12_1	464	CGVRSCPGDCRGRGRCESGR-CMCWPGYTGRDCGTRACPGDCRGRGRCVD-GRCVCNPGF
DNA 33094	854	YGVNCDKANCSTTTCFNGGTCFYPGKICICPPGLEGEQCEISKCPQPCRNGGKCIKSKCKC
		* *
HSTNX12_1	522	TGEDCGSRRCPGDCRGHGLC-EDGVCVCDAGYSGEDCSTRSCPGGCRGRGQCILD-GRCVC
DNA33094	1034	SKGYQGDLCSPVCEPGCGAHGTCHEPNKCQCQEGWHGRHCKRYEASLIH
		** *
HSTNX12_1	363	EDGYSGEDCGVRQCPNDCSQHGVQCQD-GVCICWEGYVSEDCSIRTCPNSCH

FIG. 13B-1B

34 / 108

Score = 307 (108.1 bits), Expect = 2.9e-23, P = 2.9e-23  
 Identities = 56/163 (34%), Positives = 70/163 (42%), at 674,526, Frame = +2

DNA 33094	674	CQQAECPPGGCRNGGFCNERRICECPDGFHGHCEKALCTPRCMNGGLCVTPGFICPPPGF
HSTNX12_1	526	CGSRRCPGDCRGIHGLC-EDGVCVCDAGYSGEDCSTRSCPGGCRGRGQCLD-GRCVCEGGY
DNA 33094	854	YGVNCDKANCSTTCFNGGTCFYPGKICPPGLEGEQCEISKCPQPCRNGGKICIGSKKCK
HSTNX12_1	584	SGEDCGVRQCPNDCSQHGVC-QDGVCIWEGYVSEDCSIRTCPNSNCHGRGRC-EEGRCLC
DNA33094	1034	SKGYQGDLCSPVCEPGCGAHGTCHEPNKCQCQEGWHGRHCNK
HSTNX12_1	642	DPGYTGPTCATRMCPADCRGRGRCVQ-GVCLCHVGYGGEDCGQ

FIG. 13B-2A

**Score = 237 (83.4 bits), Expect = 1.3e-15, P = 1.3e-15**  
**Identitles = 48/132 (36%), Positives = 60/132 (45%), at 674,619, Frame = +2**

DNA 33094 674 CQAECPGGCRNGGFCNERRICEPCDGFHGHCEKALCTPRCMNGGLCVTPGFCICPPGE  
\* \*\* \* \* \* \* . \* \* \* \* \*  
HSTNX12 1 619 CSIRTCPSNCHGRGRCEEGR-CLCDPGYTGTATRMCPADCRGRRCVQ-GVC LCHVG

DNA 33094	854	YGVNCDKANCSTTCFNGGT-----CFYPGKCICPPGLEGEQEISKCPQPCRNGGKCIGK
		* . * . . . * * * * * * * * * * * *
HSTNX12 1	677	GGEDCGQEEPPASACPGGCGPRELC-RAGQCVCVEGFRGPDCAIQTCPGDCRGRGEGCHDG

DNA33094	1019	SKCKCSKGYQGDLCCK * * * ** * . *
HSTN12 1	736	S-CVCKDGYAGEDCGE

**Score = 160 (56.3 bits), Expect = 3.1e-07, P = 3.1e-07**  
**Identities = 35/100 (35%), Positives = 44/100 (44%), at 671,649. Frame = +2**

DNA 33094 671 TCQQAECPGGCRNGGFCNERRICECPDGFHGHCKE-----ALCTPRCMNGGLCVTPGFC  
HSTNX12 1 649 TCATRMCPADCRGRGRCVQG-VCLCHVGYGGEDCGQEEPPASACPGGGCGPRELC-RAGQC

DNA 33094 836 ICPPGFYGVNCDKANCSTTCFNGGTCFYPGKCICPPGLEGEQC  
HSTNX12 1 707 VCVGFRGPDCAIQTCPGDCRGRGEC-HDGSVCCKDGYAGEDC

**FIG. 13B-2B**

(SEQ ID NO: 74)

Met Gly Thr	Lys	Ala	Gln	Val	Glu	Arg	Lys	Leu	Leu	Cys	Leu	Cys	Ser	Leu	Ala	Leu	Gly	Ser	Val	Thr								
1			5				10				15			20		25			30									
Val His	Ser	Ser	Glu	Pro	Glu	Val	Arg	Ile	Pro	Glu	Asn	Asn	Pro	Val	Lys	Leu	Ser	Cys	Ala	Tyr	Ser	Gly	Phe	Ser	Ser	Pro	Arg	Val
			35				40				45			50		55			60									
Glu Trp	Lys	Phe	Asp	Gln	Gly	Asp	Thr	Thr	Arg	Leu	Val	Cys	Tyr	Asn	Asn	Lys	Ile	Thr	Ala	Ser	Tyr	Glu	Asp	Arg	Val	Thr	Phe	Leu
			65				70				75			80					85								90	
Pro Thr	Gly	Ile	Thr	Phe	Lys	Ser	Val	Thr	Arg	Glu	Asp	Thr	Gly	Thr	Tyr	Thr	Cys	Met	Val	Ser	Glu	Glu	Gly	Gly	Asn	Ser	Tyr	Gly
			95				100						105					110					115				120	
Glu Val	Lys	Val	Lys	Leu	Ile	Val	Leu	Val	Pro	Pro	Ser	Lys	Pro	Thr	Val	Asn	Ile	Pro	Ser	Ser	Ala	Thr	Ile	Gly	Asn	Arg	Ala	Val
			125				130						135					140					145				150	
Leu Thr	Cys	Ser	Glu	Gln	Asp	Gly	Ser	Pro	Pro	Ser	Glu	Tyr	Thr	Trp	Phe	Lys	Asp	Gly	Ile	Val	Met	Pro	Thr	Asn	Pro	Lys	Ser	Thr
			155				160						165					170					175				180	
Arg Ala	Phe	Ser	Asn	Ser	Ser	Tyr	Val	Leu	Asn	Pro	Thr	Thr	Gly	Glu	Leu	Val	Phe	Asp	Pro	Leu	Ser	Ala	Ser	Asp	Thr	Gly	Glu	Tyr
			185				190						195					200					205				210	
Ser Cys	Glu	Ala	Arg	Asn	Gly	Tyr	Gly	Thr	Pro	Met	Thr	Ser	Asn	Ala	Val	Arg	Met	Glu	Ala	Val	Glu	Arg	Asn	Val	Gly	Val	Ile	Val
			215				220						225					230					235				240	
Ala Ala	Val	Leu	Val	Thr	Leu	Ile	Leu	Leu	Gly	Ile	Leu	Val	Phe	Gly	Ile	Trp	Phe	Ala	Tyr	Ser	Arg	Gly	His	Phe	Asp	Arg	Thr	Lys
			245				250						255					260					265				270	
Lys Gly	Thr	Ser	Ser	Lys	Lys	Val	Ile	Tyr	Ser	Gln	Pro	Ser	Ala	Arg	Ser	Glu	Gly	Glu	Phe	Lys	Gln	Thr	Ser	Ser	Phe	Leu	Val	
			275				280						285					290					295				299	

FIG. 14

37 / 108

(SEQ ID NO: 75)  
 GTCTGTCC AGGAGTCCTT CGGGGGCTGT TGTGTAGTG GCCTGATCGC GATGGGGACA AAGGGCGAAG TCGAGAGGAA ACTGTTGTGC CTCTTCATAT 100  
 TGGCGATCCT GTTGTGCTCC CTGGCATTGG GCAGTGTTAC AGTGCATCTT TCTGAACCTG AAGTCAGAAAT TCCTGAGAAT AATCCTGTGA AGTTGTCCCTG 200  
 TGCCTACTCG GGCTTTTCTT CTCCCCGTGT GGAGTGAAG TTTGACCAAG GAGACACCAC CAGACTCGTT TGCTATAATA ACAAGATCAC AGCTTCCTAT 300  
 GAGGACCGGG TGACCTTCTT GCCAACTGGT ATCACCTTCA AGTCCGTGAC ACGGGAAGAC ACTGGGACAT ACACTTGTAT GGTCTCTGAG GAAGGGGCA 400  
 ACAGCTATGG GGAGGTCAAG GTCAAGCTCA TCGTGCTTGT GCCTCCATCC AAGCCTACAG TTAACATCCC CTCCTCTGCC ACCATTGGGA ACCGGGCAGT 500  
 GCTGACATGC TCAGAACAAAG ATGGTTCCCC ACCTTCTGAA TACACCTGGT TCAAAGATGG GATAGTGATG CCTACGAATC CCAAAAGCAC CCGTGCCTTC 600  
 AGCAACTCTT CCTATGTCCT GAATCCACA ACAGGAGAGC TGGTCTTTGA TCCCCCTGCA GCCTCTGATA CTGGAGAATA CAGCTGTGAG GCACGGAATG 700  
 GGTATGGGAC ACCCATGACT TCAATGCTG TGGCGATGGA AGCTGTGGAG CGGAATGTGG GGGTCATCGT GGCAGCCGTC CTTGTAACCC TGATTCCTCT 800  
 GGGAACTTG GTTTTGGCA TCTGTTTGC CTATAGCCGA GGCCACTTTG ACAGAACAAA GAAAGGAGT TCGAGTAAGA AGGTGATTTA CAGCCAGCCT 900  
 AGTGCCCGAA GTGAAGGAGA ATTCRAACAG ACCTCGTCAAT TCCTGGTGTG AGCCTGGTGG GCTCACCCGC TATCATCTGC ATTTGCCCTTA CTCAGGTGCT 1000  
 ACCGGACTCT GGCCCTGAT GTCTGTAGTT TCACAGGATG CCTTATTGT CTCTACACC CCACAGGGCC CCTACTTCT TCGGATGTGT TTTTAATAAT 1100  
 GTCAGCTATG TGCCCCATCC TCCTTCATGC CCTCCCTCCC TTTCCCTACCA CTGCTGAGTG GCCTGGAAT TGTTTAAAGT GTTTATTCCC CATTTCTTTG 1200  
 AGGATCAGG AAGGATCCT GGGTATGCCA TTGACTTCCC TTCTAAGTAG ACAGCAAAA TGGCGGGGGT CGCAGGAATC TGCACCTCAAC TGCCCACTG 1300  
 GCTGGCAGGG ATCTTTGAAT AGGTATCTTG AGCTTGGTTC TGGGCTCTTT CCTTGTGTAC TGACGACCAG GSCCAGCTGT TCTAGAGCGG GAATTAGAGG 1400  
 CTAGAGGGC TGAATGGTT GTTTGGTGAT GACACTGGGG TCCTTCCATC TCTGGGGCCC ACTCTCTTCT GTCTTCCCAT GGGAAAGTGCC ACTGGGATCC 1500  
 CTCTGCCCCG TCCTCTCTGA TACAAGCTGA CTGTGTCTGT GGAAAATGGG AGCTCTTGTG GTGAGAGCA TAGTAAATTT TCAGAGAACT 1600  
 TGAAGCCCAA AGGATTTAAA ACCGCTGCTC TAAAGAAAAA AAAACTGGAG GCTGGGGCCA GTGGCTCAG CCTGTAAATCC CAGAGGCTGA GGCAGGCGGA 1700  
 TCACCTGAGG TCGGGAGTTC GGGATCAGCC TGACCAACAT GGAGAAACCC TACTGGAAT ACAAGTTAG CCAGGCATGG TGGTGATGC CTGTAGTCCC 1800  
 AGCTGCTCAG GAGCCTGGCA ACNAGAGCAA AACTCCAGCT CA 1842

FIG. 15

38 / 108

**Consen0870: 4 members (3 incyte, 1 est) 390 bp, 0 gaps, 153 bp orf (+3)**

1452523	1	CTTCTTGCCAACTGGTATCACCTTCAAGTCCGTGACACGGGAAGACACTG
SEQ ID NO:15		
2345419	1	CACGGGAAGACACTG
SEQ ID NO:16		
<DNA35936>	1	CTTCTTGCCAACTGGTATCACCTTCAAGTCCGTGACACGGGAAGACACTG
SEQ ID NO:3		
1452523	51	GGACATACTTGTATGGTCTCTGAGGAAGGCGGCAACAGCTATGGGGAG
2345419	16	GGACATACTTGTATGGTCTCTGAGGAAGGCGGCAACAGCTATGGGGAG
T87045	1	GAG
SEQ ID NO:17		
<DNA35936>	51	GGACATACTTGTATGGTCTCTGAGGAAGGCGGCAACAGCTATGGGGAG
1452523	101	GTCAAGGTCAAGCTCATCGTGCTTGTGCCTCCATCCAAGCCTACAGTTAA
2345419	66	GTCAAGGTCAAGCTCATCGTGCTTGTGCCTCCATCCAAGCCTACAGTTAA
T87045	4	GTCAAGGTCAAGCTCATCGTGCTTGTGCCTCCATCCAAGCCTACAGTTAA
<DNA35936>	101	GTCAAGGTCAAGCTCATCGTGCTTGTGCCTCCATCCAAGCCTACAGTTAA
1452523	151	CATCCCCTCCTCTGCCACCATTGGGAACCGGGCAGTGCTGACATGCTCAG
2345419	116	CATCCCCTCCTCTGCCACCATTGGGAACCGGGCAGTGCTGACATGCTCAG
T87045	54	CATCCCCTCCTCTGCCACCATTNGGAACCGGGCAGTGCTGACATGCTCAG
1508565	1	TTGGGAACCGGGCAGTGCTGACATGCTCAG
SEQ ID NO:18		
<DNA35936>	151	CATCCCCTCCTCTGCCACCATTGGGAACCGGGCAGTGCTGACATGCTCAG
1452523	201	AACAAGATGGTTCCCCACCTTCTGAATACACCTGGTTCAAAGATGGG
2345419	166	AACAAGATGGTTCCCCACCTTCTGAATACACCTGGTTCAAAGATGGGATA
T87045	104	AACAAGATGGTTCCCCACCTTCTGAATACACCTGGTTCAAAGATGGGATA
1508565	31	AACAAGATGGTTCCCCACCTTCTGAATACACCTGGTTCAAAGATGGGATA
<DNA35936>	201	AACAAGATGGTTCCCCACCTTCTGAATACACCTGGTTCAAAGATGGGATA
2345419	216	GTGATGCCTACGAATCCCAAAGCACCCGTGCCTT
T87045	154	GTGATGCCTACGAATCCCAAAGCACCCGTGCCTTCAGCAACTCTTCCTA
1508565	81	GTGATGCCTACGAATCCCAAAGCACCCGTGCCTTCAGCAACTCTTCCTA
<DNA35936>	251	GTGATGCCTACGAATCCCAAAGCACCCGTGCCTTCAGCAACTCTTCCTA
T87045	204	TGTCCTGAATCCCAACAGGAGAGCTGGTCTTTGATCCCCTGTCAGCCT
1508565	131	TGTCCTGAATCCCAACAGGAGAGCTGGTCTTTGATCCCCTGTCAGCCT
<DNA35936>	301	TGTCCTGAATCCCAACAGGAGAGCTGGTCTTTGATCCCCTGTCAGCCT
T87045	254	CTGATACTNGAGAATACAGCTGTGAGGCACGGAATGGGTA
1508565	181	CTGATACTGGAGAATACAGCTGTGAGGCACGGAATGGGTA
<DNA35936>	351	CTGATACTGGAGAATACAGCTGTGAGGCACGGAATGGGTA

**FIG. 16**

39 / 108

(SEQ ID NO: 92) C17760  
 (SEQ ID NO: 93) W76302  
 (SEQ ID NO: 94) 3124762  
 (SEQ ID NO: 95) AA215609  
 (SEQ ID NO: 96) 777818  
 (SEQ ID NO: 97) 3234064  
 (SEQ ID NO: 98) 1298110  
 (SEQ ID NO: 99) AA101519  
 (SEQ ID NO: 100) 2197534  
 (SEQ ID NO: 101) AA101561  
 (SEQ ID NO: 102) AA227408  
 (SEQ ID NO: 103) 2612024  
 (SEQ ID NO: 104) 492141  
 (SEQ ID NO: 105) 2252527  
  
 (SEQ ID NO: 77) <consen01>

C17760  
 W76302  
 3124762  
 AA215609  
 777818  
 3234064  
 1298110  
 AA101519  
 2197534  
 AA101561  
 AA227408  
 2612024  
 492141  
 2252527  
 (SEQ ID NO: 106) 2456003  
 (SEQ ID NO: 107) 2861301  
 (SEQ ID NO: 108) 3236257

&lt;consen01&gt;

C17760  
 W76302  
 3124762  
 AA215609  
 777818  
 3234064  
 1298110  
 AA101519  
 2197534  
 AA101561  
 AA227408  
 2612024  
 492141  
 2252527  
 2456003  
 2861301  
 3236257  
 (SEQ ID NO: 109) 014756  
  
 <consen01>

1 TCTCAGTCCCCTCGCTGTAGTCGCGGAGCTGTGTTCTGTTTCCCAGGAGT  
 1 CGTAGTCGCGGNNNGTNGT-CTGTT-CCCAGGAGT  
 1 GTCGCGGAN-TGTGT-CTGTT-CCCAGGAGT  
 1 CGCGGNGTGTGT-CTGTT-CCCAGGAGT  
 1 CGCGGAGCTGTGT-CTGTT-CCCAGGAGT  
 1 GGANTGTTGTCTGTT-CCCAGGAGT  
 1 GCTGTGT-CTGTT-CCCAGGAGT  
 1 TCTGTT-CCCAGGAGT  
 1 GTCTGTT-CTCAGGAGT  
 1 TCTGTT-CCCAGGAGT  
 1 TCTGTT-CCCAGGAGT  
 1 CTCAGGAGT  
 1 CAGGAGT  
 1 AGGAGT  
 1 ++++++  
 1 TCTCAGTCCCCTCGCTGTAGTCGCGGAGCTGTGTTCTGTTTCCCAGGAGT

51 CCTTCGGCGG-C-TGTTGTGCTCAGGTGCGCCTGATCGCGATGGGGCACA  
 34 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 29 CCTTCGGCGG-C-TGTNGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 28 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 28 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 25 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 22 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 16 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 17 CCTTCGGCGG-C-TGTTGTG-TC-GG-GAGCCTGATCGCGATGGGG-ACA  
 16 CCTTCGGCGG-CATGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 16 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 10 CCTTCGGCGG-C-TGTTGTG-TC-GG-GAGCCTGATCGCGATGGGG-ACA  
 8 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 7 CCTTCGGCGG-C-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 1 CCTTCGGCGGNC-TGTTGTG-TCAG-TG-GCCTGATCGCGATGGGG-ACA  
 1 CGG-C-TGTTGTG-TC-GG-GAGCCTGATCGCGATGGGG-ACA  
 1 GTCAGTGGCCTGATCGCGATGGGG-ACA  
 ++++++ + +++.+++++ ..... ++++++ ++++++  
 51 CCTTCGGCGG C TGTTGTGCTCAGGTGCGCCTGATCGCGATGGGG ACA

99 AAGGCGCAAGCTCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 78 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 73 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 72 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 72 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 69 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 66 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 60 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 61 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 61 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 60 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 54 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCANATTGGCGANCC  
 52 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTNTTCATATTGGCGATCC  
 51 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 46 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 38 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 28 AAGGCGCAAG-TCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC  
 1 GGAGGAAGCATCTGGCTGGCAGGAAGTGGGTGCTGGGCC  
 ++++++ ..... ++++++ ++++++ ++++++  
 98 AAGGCGCAAGCTCGAGAGGAACTGTTGTGCCTCTTCATATTGGCGATCC

FIG. 17A





C17760		292	AGACACCACCAG
W76302		269	AGACACCACCAGACTCGTTTGCTATAATAACAAGATCACAGCTTCCTATG
3124762		264	AGACACCA
777818		263	AGACAC
3234064		259	A
1298110		257	A
2197534		252	AGACA
AA101561		252	AGACACCACCAGACTCGTTTGCTATAATAACAAGATCACAGCTTCCTATG
492141		243	AGACAACACCAGACT
2252527		242	AGACACCACCA
2861301		229	CACCCAGCCCAGCCTGCAGTTTGGGGGTGTTCCCCATCTCGTGTAT
3236257		219	AGACACCACCAGACTCGTTTGCTATAATAACAAGATCACAGC
014756		182	AGACACCACCAGACTCGTTTGCTATAATAACAAGATCACAGCTTCCTATG ...+...+++++.++....+.....+.....+...+..++++
<consen01>		291	AGACACCACCAGACTCGTTTGCTATAATAACAAGATCACAGCTTCCTATG
W76302		319	AGGACCGGGTGACCTTCTTGCCAACCTGG-TATCACCTTC-AAGTCCGTG
AA101561		302	AGGACCGGGTGACCTTCTTGCCAACCT-GGTATCACCTTC-AAGTCCGTG
014756		232	AGGACCGGGTGACCTTCTTGCCAACCT-GG-TATCACCTTCNAAGTNCGTG
(SEQ ID NO: 76) DNA35936.init		1	CTTCTTGCCAACCT-GG-TATCACCTTC-AAGTCCGTG
(SEQ ID NO: 110) 1452523		1	CTTCTTGCCAACCT-GG-TATCACCTTC-AAGTCCGTG
(SEQ ID NO: 111) T73746		1	TC-ANACNNTN +++++++ ++++++ ++ ++++++ +...+.+
<consen01>		341	AGGACCGGGTGACCTTCTTGCCAACCT GG TATCACCTTC AAGTCCGTG
W76302		367	ACACGGGAAGACACT-GGGACATACACTT
AA101561		350	ACACNGGAAGACACT-GGGACATACACTTGTATGGTCTCTGAGGAAGGC
014756		280	ACACGGGAAGACACT-GGGACATACACTTGTATG
DNA35936.init		35	ACACGGGAAGACACT-GGGACATACACTTGTATGGTCTCTGAGGAAGGC
1452523		35	ACACGGGAAGACACT-GGGACATACACTTGTATGGTCTCTGAGGAAGGC
T73746		12	ACACNGGAAGACACTTGGGNATACACTTGTATGGACTCTNAGGANNGC
(SEQ ID NO: 112) 2345419		1	CACGGGAAGACACT-GGGACATACACTTGTATGGTCTCTGAGGAAGGC ++++:++++ ++++++ +++..+++++. ....+.++++.++++.++
<consen01>		388	ACACGGGAAG GACACT GGGACATACACTTGTATGGTCTCTGAGGAAGGC
AA101561		399	GGCAACAGCTATGGGGA
DNA35936.init		83	GGCAACAGCTATGGGGAGGTCAAGGTCAAGCTCATCGTGCTTGTGCCCTCC
1452523		83	GGCAACAGCTATGGGGAGGTCAAGGTCAAGCTCATCGTGCTTGTGCCCTCC
T73746		61	GGCAACAGCTATGGGNNNGNCAAGGTCANNCTCATCNTNCTTCNNCCCTGC
2345419		48	GGCAACAGCTATGGGGAGGTCAAGGTCAAGCTCATCGTGCTTGTGCCCTCC
(SEQ ID NO: 113) 1731885		1	CTATGGGGAGGTCAAGGTCAAGCTCATCGTGCTTGTGCCCTCC
(SEQ ID NO: 114) T84016		1	AGAGGTCAAGGTCAAGCTCATCGTGCTTGTGCCCTCC
(SEQ ID NO: 115) T87045		1	GAGGTCAAGGTCAAGCTCATCGTGCTTGTGCCCTCC
(SEQ ID NO: 116) 1932979		1	GTCAGGTCAAGCTCATCGTGCTTGTGCCCTCC +++++++ +...+.+++++. ++++++. ++++++. +...+. +...+. +
<consen01>		436	GGCAACAGCTATGGGGAGGTCAAGGTCAAGCTCATCGTGCTTGTGCCCTCC
DNA35936.init		133	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG
1452523		133	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG
T73746		111	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCNCATTGGGNACCGGG
2345419		98	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG
1731885		43	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG
T84016		37	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG
T87045		36	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG
1932979		27	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG
(SEQ ID NO: 117) 1508565		1	TTGGGAACCGGG
(SEQ ID NO: 118) 1508552		1	TTGGGAACCGGG
(SEQ ID NO: 119) R02633		1	G +++++++ +...+.+++++. ++++++. ++++++. +...+. +...+. +
<consen01>		486	ATCCAAGCCTACAGTTAACATCCCCTCCTCTGCCACCATTGGGAACCGGG

FIG. 17C

42 / 108

DNA35936.init	183	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
1452523	183	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
T73746	161	CAGTGTGANATGCTCAGAACACGNTGGTTCCCCACCT
2345419	148	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
1731885	93	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
T84016	87	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
T87045	86	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
1932979	77	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
1508565	13	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
1508552	13	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
R02633	2	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
		+++++.+++++.+++++.+++++.+++++.+++++.+++++.+++++.+++++
<consen01>	536	CAGTGTGACATGCTCAGAACAAGATGGTTCCCCACCTTCTGAATACACC
DNA35936.init	233	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
1452523	233	TGGTTCAAAGATGGG
2345419	198	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
1731885	143	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
T84016	137	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
T87045	136	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
1932979	127	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
1508565	63	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
1508552	63	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
R02633	52	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
(SEQ ID NO: 120) 979636	1	GGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
		+++++.+++++.+++++.+++++.+++++.+++++.+++++.+++++.+++++
<consen01>	586	TGGTTCAAAGATGGGATAGTGATGCCTACGAATCCCAAAGCACCCTGCG
DNA35936.init	283	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGAGCTGG
2345419	248	CTT
1731885	193	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGAGCTGG
T84016	187	CTT-CAGCAACTCTTCCTATGTCCTGGAATCCCACAACAGGGAGAGCTGG
T87045	186	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGAGCTGG
1932979	177	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGANCTGG
1508565	113	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGAGCTGG
1508552	113	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGAGCTGG
R02633	102	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGAGCTGG
979636	39	CTT-CAGCAACTCTTCCTATGTCCTG-AATCCCACAACAGG-AGAGCTGG
(SEQ ID NO: 121) AA404390	1	AGCTGG
(SEQ ID NO: 122) 2328920	1	GG
		+++ ++++++.+++++.+++++.+++++.+++++.+++++.+++++.+++++
<consen01>	636	CTT CAGCAACTCTTCCTATGTCCTG AATCCCACAACAGG AGAGCTGG
DNA35936.init	330	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACTGG-AG-AATACAGCTGT
1731885	240	-TCTTT-GATCCCCTGT-CAGCCTC
T84016	237	GTCTTTTGATCCCCTGTT-CAGCCTCTGGATANTGGGAGGANTACAGCTGT
T87045	233	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACTNG-AG-AATACAGCTGT
1932979	224	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACT
1508565	160	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACTGG-AG-AATACAGCTGT
1508552	160	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACTGG-A
R02633	149	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACTGGGAG-AATACAGCTGT
979636	86	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACTGG-AG-AATACAGCTGT
AA404390	7	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACTGG-AG-AATACAGCTGT
2328920	3	-TCTTT-GATCCCCTGT-CAGCCTCTG-ATACNGG-AG-AATACAGCTGT
(SEQ ID NO: 123) 2925803	1	CTGT
		+++++ ++++.+++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++
<consen01>	683	TCTTT GATCCCCTGT CAGCCTCTG ATACTGG AG AATACAGCTGT

FIG. 17D

[illegible]

FIG. 17E

R02633		381	AAAGGAAAGGGGANTTTCGATTAAAGGAAGGTGNTTTACAGCCAGCCTACTT
979636		300	AAAG-AAAGGGGACTT-CGAGTAAG-A
2328920		217	AAAG-AAAGGGGACTT-CGAGTAAG-AAGGTGATTTACAGCCAGCCTAGTG
2925803		175	AAAG-AAAGGGGACTT-CGAGTAAG-AAGGTGATTTACAGCCAGCCTAGTG
1519947		164	GTAT-CTGCCCCCAG-AGGCTCTC-CTTTGTACTGCC
1521745		164	GTAT-CTGCCCCCAG-AGGCTCTC-CTTTGTACTGCCCCCATCC
AA152150		164	AAAG-AAAGGGGACTT-CGAGTAAG-AAGGTGATTTACAGCCAGCCTAGTG
1610836		98	AAAG-AAAGGGGACTT-CGAGTAAG-AAGGTGATTTACAGCCAGCCTAGTG
1274809		68	AAAG-AAAGGGGACTT-CGAGTAAG-AAGGTGATTTACAGCCAGCCTAGTG
956595		52	AAAG-AAAGGGGACTT-CGAGTAAG-AAGGTGATTTACAGCCAGCCTAGTG
1818676		23	AAAG-AAAGGGGACTT-CGAGTAAG-AAGGTGATTTACAGCCAGCCTAGTG
(SEQ ID NO: 131)	2220993	1	GTGATTTACAGCCAGCCTAGTG
(SEQ ID NO: 132)	1706515	1	GATTTCAGCCAGCCTAGTG
(SEQ ID NO: 133)	N28398	1	AGCCTAGTG
(SEQ ID NO: 134)	360948	1	CCCGTCGCNC
	<consen01>		...+. .... .+..+... ..+..+..+..+..+..+
		897	AAAG AAAGGGGACTT CGAGTAAG AAGGTGATTTACAGCCAGCCTAGTG
<hr/>			
R02633		431	
2328920		264	CCCGAA
2925803		222	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
AA152150		211	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
1610836		145	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
1274809		115	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
956595		99	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
1818676		70	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
2220993		23	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGNGTGAGCC
1706515		21	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
N28398		10	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
360948		10	CCCGAAGTGAAGGAGAATTCAAACAGNCCCTCGTCATTCTGGNGTGAGCC
(SEQ ID NO: 135)	3240004	1	GTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
(SEQ ID NO: 136)	2044611.RC	1	CGGCTCGAGCGTCATTCTGGTGTGAGCC
	<consen01>		+.....+.....+.....+.....+.....+
		944	CCCGAAGTGAAGGAGAATTCAAACAGACCTCGTCATTCTGGTGTGAGCC
<hr/>			
2925803		272	TGGTCGGCTC
AA152150		261	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACTG
1610836		195	TGGT
1274809		165	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG
956595		149	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG
1818676		120	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG
2220993		73	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG
1706515		71	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG
N28398		60	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACTG
360948		60	TGGNCGGN TNACC GNCTATCATCTGCATTGTCCT TACTNAGGTGN TACCG
3240004		45	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG
2044611.RC		30	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG
(SEQ ID NO: 137)	2382718	1	CTGCATTGTCCTTACTCAGGTGCTACCG
	<consen01>		+++..++..+..++++..+++++++.....++++..++++..+++.
		994	TGGTCGGCTCACCGCCTATCATCTGCATTGTCCTTACTCAGGTGCTACCG

FIG. 17F

	AA152150	311	GACTCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	1274809	215	GACTCTGGCCCCCTG-ATGTCTGTGA
	956595	199	GACTCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	1818676	170	GACTCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	2220993	123	GACTCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	1706515	121	GACTCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	N28398	110	GACTCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	360948	110	GACTNTGGNCCCTG-ATGTCTGTAGTTTCANAGGNTGCCTTATTTGTCTT
	3240004	95	GACTCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	2044611.RC	80	GACTCTGGCCCCCTG-ATGTCTGTGA-TTTCACAGGATGCCTTATTTGTCTT
	2382718	29	GACTCTGGCCCCCTGGATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
(SEQ ID NO: 138)	R28222	1	CTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
(SEQ ID NO: 139)	1889866	1	TCTGGCCCCCTG-ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
(SEQ ID NO: 140)	T39607	1	TAGTTTCACAGGATGCCTTATTTGTCTT
(SEQ ID NO: 141)	T39606	1	AGTTT-ACAGGAT-CCTTATTTGTCT-
(SEQ ID NO: 142)	1424836	1	GGATGCCTTATTTGTCTT
(SEQ ID NO: 143)	AA224590.RC	1	TAACAA
(SEQ ID NO: 144)	929944	1	T
(SEQ ID NO: 145)	930239	1	T
			++++.+++..+++++ ++++++.+++..+++++.++..
<consen01>		1044	GACTCTGGCCCCCTG ATGTCTGTAGTTTCACAGGATGCCTTATTTGTCTT
	AA152150	360	CTACACCCCACAGGGCCCCC-TACTTCTN
	956595	248	CTACACCCCACAGGGCCCCC-TACTTCTT
	1818676	219	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	2220993	172	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	1706515	170	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	N28398	159	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	360948	159	CTACAACCCACAGGGNCCCC-TACTTCTTCGG-A-TGTGTTTTT-AA
	3240004	144	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	2044611.RC	128	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	2382718	79	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	R28222	46	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	1889866	47	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	T39607	29	CTACACCCCACAGGGCCCCCGTACTTCTTCGGNA-TGTGTTTTT-AAATAA
	T39606	25	CTACACCC-ACAGG-CCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	1424836	19	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
	AA224590.RC	7	CCCCAACAGGGNCCCCCNTA-ACCTTCTTCGN-AATGTGTTTTTAAATAA
	929944	2	CTACACCCCACAGGGCCCCC-TACTTNTTCGG-A-TGTGTTTTT-AAATAA
	930239	2	CTACACCCCACAGGGCCCCC-TACTTCTTCGG-A-TGTGTTTTT-AAATAA
(SEQ ID NO: 146)	876764	1	CTTCGG-A-TGTGTTTTT-AAATAA
(SEQ ID NO: 147)	159097	1	ACGGA-A-TGTGTTTTT-AAATAA
			+...+.~.....+.~ ..+++.~.+ + ~+..++++ ++...
<consen01>		1093	CTACACCCCACAGGGCCCCC TACTTCTTCGG A-TGTGTTTTT-AAATAA

FIG. 17G

46 / 108

1818676	265	TGTC-AGCTATGTGCCCC--ATCCTCCTT-C
2220993	218	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CA
1706515	216	TGTC-AGCTA
N28398	205	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
3240004	190	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
2044611.RC	174	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCNTN
2382718	125	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
R28222	92	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
1889866	93	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
T39607	77	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
T39606	69	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
1424836	65	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
AA224590.RC	55	TGTCAGCTATGTGCCCCCAATCCTCCTTTCATGCCCCCTCCCTTCCCTT
929944	48	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
930239	48	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
876764	22	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
159097	21	TGTC-AGCTATGTGCCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
(SEQ ID NO: 148) 1004380	1	CCCC--ATCCTCCTT-CATGCCC--TCCCT-CCCTT
(SEQ ID NO: 149) 1217411	1	TGCCC--TCCCT-CCCTT
<consen01>		++++ ++++++ ++.+++++ ++++.+ +.+++ ++.+. .
	1139	TGTC AGCTATGTGCCCC ATCCTCCTT CATGCCC TCCCT CCCTT
N28398	248	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
3240004	233	TCC-TACCA
2044611.RC	217	NCC-TACCACTGCTGAGTGGC
2382718	168	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
R28222	135	TCC-TACCACTGCTGAGTGGCC-TGGAA-CTTGTTTAAAGTGTTATTTC
1889866	136	TCC-TACCA-TGCTGAGTGGCC-TGG-AA--TTGTTNAGGNGTTAATNC
T39607	120	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
T39606	112	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
1424836	108	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
AA224590.RC	105	TCCCTACCACTGCTGAGTGGCCCTGG-AACTTGTTTAAAGTGTTATTTC
929944	91	TNC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
930239	91	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
876764	65	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
159097	64	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
1004380	32	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
1217411	16	TCC-TACCACTGCTGAGTGGCC-TGG-AA-CTTGTTTAAAGTGTTATTTC
(SEQ ID NO: 150) AA483522.RC	1	C-TGG-AA-CTTGTTTAAAGTGTTATTTC
(SEQ ID NO: 151) 732999	1	GTTTAAAGTGTTATTTC
<consen01>		..+ ++++++ +++ ++ ++++++.+.+.+.+.+. .
	1182	TCC TACCACTGCTGAGTGGCC TGG AA CTTTGTTTAAAGTGTTATTTC

FIG. 17H

N28398	294	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
2382718	214	CCC--ATT
R28222	182	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
1889866	180	CCC--AATTNNTGGGGGTTAGGAAA
T39607	166	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
T39606	158	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
1424836	154	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
AA224590.RC	154	CCCCAATTTCTTTGAGGGATCANGAAGGAATCCTGGGTATGCC-AATTGA
929944	137	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
930239	137	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
876764	111	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
159097	110	CCC--ATTTNNTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
1004380	78	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
1217411	62	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
AA483522.RC	27	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
732999	18	CCC--ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
(SEQ ID NO: 152)	1	ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
(SEQ ID NO: 153)	1	ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC-A-TTGA
(SEQ ID NO: 154)	1	TCAGGAAGGAATCCTGGGTATGCCA-TTGA
(SEQ ID NO: 155)	1	GGAAGGAATCCTGGGTATGCC-A-TTGA
(SEQ ID NO: 156)	1	AAGGAATCCTGGGTATGCC-A-TTGA
(SEQ ID NO: 157)	1	AGGAATCCTGGGTATGCC-A-TTGA
(SEQ ID NO: 158)	1	ATGCC-A-TTGA
(SEQ ID NO: 159)	1	GCC-A-TTGA
(SEQ ID NO: 160)	1	GCC-A-TTGA
	+++	+.++..+++..++++..+++++++..+++++++ + +++++
<consen01>	1228	CCC ATTTCTTTGAGGGATCAGGAAGGAATCCTGGGTATGCC A TTGA
N28398	340	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGGGTC-GCCAGG-A
R28222	228	CTTCCCTTCTAA-GTAGGACAGCAAAAATGGGCGGGG-TC-GC-AGGGA
T39607	212	CTTCCCTTCTAA-GTAG-ACAGCAA
T39606	204	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1424836	200	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
AA224590.RC	203	CTTCCCTTCTAAAGTAG-ACAGCAAAAATGG-CGGGGGGTC-GC-AGG-A
929944	183	CTTCCCTTCTAA-GT-G-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
930239	183	CTTCCNTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
876764	157	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
159097	156	NTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1004380	124	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1217411	108	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
AA483522.RC	73	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-CC-AGG-A
732999	64	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1282058	44	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1283885	44	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
N20044.RC	31	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TCGGC-AGG-A
2797137	27	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
2025350	25	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
3212856	24	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1611708	11	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1807742	9	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
1804959	9	CTTCCCTTCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
(SEQ ID NO: 161)	1	TGGACCTAA-GTAG-ACAGCAAAAATGG-CGGGGG-TC-GC-AGG-A
(SEQ ID NO: 162)	1	CG-GCTCGAGAAATGG-CGGGGG-TC-GC-AGG-A
(SEQ ID NO: 163)	1	GCAAAAATGG-CGGGGG-TC-GC-AGG-A
(SEQ ID NO: 164)	1	AAAATGG-CGGGGG-TC-GC-AGG-A
(SEQ ID NO: 165)	1	G-CGGGGG-TC-GC-AGG-A
(SEQ ID NO: 166)	1	A-AGG-A
(SEQ ID NO: 167)	1	C-GC-AGG-A
(SEQ ID NO: 168)	1	GG-A
	..++..+++ ++.+.+.+.+++++ +++++ ++ ..++.	
<consen01>	1274	CTTCCCTTCTAA GTAG ACAGCAAAAATGG CGGGGG TC GC AGG A

FIG. 17I



48/108

N28398	385 ATCCTGCA-CTCAA-CTGNCCCACCTTGGCTGGGCAGGGNA-TCTTTG-A
R28222	274 ATC-TGCA-CTCAA-CTG-CCCACCT-GGGTGGGCAGGGGA-TCTTTGGA
T39606	247 ATC-TGCA-CTCAA-CTG-CCC
1424836	243 ATC-TGCA-CTCAA-CTG
AA224590.RC	248 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
929944	225 ATC-TGCA-CTNA
930239	226 ATC-TGCA-CTNAA-CTG-CCCACCT-GGNTGG-CAGGG-A-TCTTTG-A
876764	200 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
159097	199 ATC-TN-A-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-N-TCTTTG-A
1004380	167 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1217411	151 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
AA483522.RC	116 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
732999	107 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1282058	87 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1283885	87 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
N20044.RC	75 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
2797137	70 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
2025350	68 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
3212856	67 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1611708	54 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1807742	52 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1804959	52 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
AA244075	41 ATC-TG-A-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-ACTCTTTGTA
1684149	30 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1793273	25 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
1345563	22 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
T40695.RC	16 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
R27969.RC	6 ATC-TGNAACTNAA-CTG-CCCCCCT-GGCTGG-CAGGGGA-TCTTNA-A
3144865	8 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
R72982	4 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
(SEQ ID NO: 169) 1752577	1 ATC-TGCA-CTCAA-CTG-CCCACCT-GGCTGG-CAGGG-A-TCTTTG-A
(SEQ ID NO: 170) T86963.RC	1 T-A
	+++ +..+ ++.++ +++ ++.+++ ++.+++ +++++ . +++++. +
<consen01>	1317 ATC TGCA CTCAA CTG CCCACCT GGCTGG CAGGG A TCTTTG A

FIG. 17J

N28398	431	AATAAG-G-TATC-TTTGG-A-GGC-TTG-G-TTC-TGGG-GCT-CCT
R28222	318	A-TA-G-GGTATC-TTT-G-A-G-C-TTG-GGTTT-TGGGGCTC-TTTTC
AA224590.RC	290	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
930239	267	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-T
876764	241	A-TA-G-G-TATC-TT--G
159097	239	A
1004380	208	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1217411	192	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
AA483522.RC	157	A-TA-G-G-TATC-TT--G-AAG-C-TTG-G-TTC-TGGG-CTC-TTT-C
732999	148	A-TA-G-GGTATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1282058	128	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1283885	128	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
N20044.RC	116	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
2797137	111	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
2025350	109	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
3212856	108	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1611708	95	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1807742	93	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1804959	93	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
AA244075	83	A-TA-GCG-TATCGTT--GTA-G-CGTTGAG-TTCGTGGG-CTCGTTT-C
1684149	71	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1793273	66	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1345563	63	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
T40695.RC	57	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
R27969.RC	49	A-AN-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGGGCTC-TTT-C
3144865	49	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
R72982	45	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
1752577	42	A-TA-G-G-TATC-TT--G-A-G-C-TTG-G-TTC-TGGGGCTC-TTN-C
T86963.RC	3	A-TA-G-G-TACC-TT--G-A-A-CNTTG-G-TCC-TGG--CNC-TTC-C
(SEQ ID NO: 171) 767739	1	T--G-A-G-C-TTG-G-TTC-TGGG-CTC-TTT-C
(SEQ ID NO: 172) 647074	1	GG-CTC-TTT-C
<consen01>	1358	A TA G G TATC TT G A G C TTG G TTC TGGG CTC TTT C
R28222	356	-CTTG-T
AA224590.RC	323	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
1004380	241	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGT
1217411	225	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-T
AA483522.RC	191	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
732999	182	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
1282058	161	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGT-GGG
1283885	161	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGT-GGG
N20044.RC	149	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGT-GGG
2797137	144	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
2025350	142	-CTTG-TG-TAC-TGACG-ACC-AGGGGCC-AGCTGTTC-TAGAGC-GGG
3212856	141	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
1611708	128	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
1807742	126	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
1804959	126	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
AA244075	123	GCTTGGTGCTACGTGACGACGAGGTCGAGCTGTTCGTTAGAGCCGGG
1684149	104	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
1793273	99	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGT-GGG
1345563	96	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
T40695.RC	90	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGT-GGG
R27969.RC	83	-CTTG-TG-TAC-TGACG-ACC-CGGG-CC-AGCTGTTC-TAGAGT-GGG
3144865	82	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
R72982	78	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
1752577	76	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-A
T86963.RC	36	-CTTG-TG-TAC-TGACG-ACCCAGGG-CCCAGCTGTTC-TAAAGC-GGG
767739	24	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
647074	10	-CTTG-TG-TAC-TGACG-ACC-AGGG-CC-AGCTGTTC-TAGAGC-GGG
(SEQ ID NO: 173) AA244018.RC	1	CC-AGCTGTTC-TAGAGC-GGG
<consen01>	1391	CTTG TG TAC TGACG ACC AGGG CC AGCTGTTC TAGAGC GGG

FIG. 17K

50 / 108

```
AA224590.RC      363 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
AA483522.RC      232 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
732999           222 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGG
1282058          201 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATG
1283885          201 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
N20044.RC        189 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
2797137          184 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
2025350          183 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
3212856          181 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
1611708          168 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGT
1807742          166 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
1804959          166 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
AA244075         173 TAATTAGAAGCCTAGAGC-GGCCTGAAAAATGTTGTNTTGGTGATGAC
1684149          144 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
1793273          139 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
1345563          136 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
T40695.RC        130 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
R27969.RC        123 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAN
3144865          122 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
R72982           118 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
T86963.RC        78 -AATTANA-GGC-TAGAAC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
767739           64 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
647074           50 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGACGAC
AA244018.RC      20 -AATTAGA-GGC-TAGAGC-GGC-TGAAA--TGTTGT-TTGGTGATGAC
(SEQ ID NO: 174) AA149993.RC      1 TTAGAAGNC-TAGAGCCGGC-TGAAA--TGTTGT-TTGGTGATGAC
(SEQ ID NO: 175) AA101562.RC      1 TGGTGATGAC
(SEQ ID NO: 176) 2223391          1 TGAC
(SEQ ID NO: 177) 1447744          1 TGAC
          +++++.+ +. +++++.+ +++ ++++++ ++++++ .+++++.+.
<consen01> 1431 AATTAGA GGC TAGAGC GGC TGAAA TGTTGT TTGGTGATGAC

AA224590.RC      405 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-NT
AA483522.RC      274 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
1283885          243 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
N20044.RC        231 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
2797137          226 A
2025350          225 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
3212856          223 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
1807742          208 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
1804959          208 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
AA244075         222 ACCTGGGGGTCTCTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
1684149          186 AC-TGGGG-TCCTTCC-ATCTCT
1793273          181 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
1345563          178 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
T40695.RC        172 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
R27969.RC        165 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
3144865          164 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
R72982           160 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
T86963.RC        120 AC-TGGGG-TCCTTCCCATCTCTGGGGCCCCACT-CTCTT-C-TGTCCTT
767739           106 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
647074           92 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
AA244018.RC      62 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
AA149993.RC      43 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
AA101562.RC      11 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
2223391          5 AC-TGNGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
1447744          5 AC-TGGGG-TCCTTCC-ATCTCTGGGGCCCC-ACT-CTCTT-C-TGTC-TT
(SEQ ID NO: 178) R01692.RC        1 GGGGN-CCCTTCCCATCTCTGGGGCCCCACTNCTCCT-CCTGTCCTT
(SEQ ID NO: 179) R87078           1 TAAAGACGGAGCT-CGC-TCTGT-C-ACCC-AG
(SEQ ID NO: 180) AA101520.RC      1 ATCTCTGGGGCCCC-ACT-CTCNT-C-TGTC-TT
(SEQ ID NO: 181) T84017.RC        1 CTTN-C-TGTC-CT
(SEQ ID NO: 182) 1208791          1 CTCTT-C-TGTC-TT
(SEQ ID NO: 183) 1208826          1 CTCTT-C-TGTC-TT
(SEQ ID NO: 184) 143613          1 CTCTN-T-TGTC-TT
          ++ .+.+. .+++++ .+++++ .+.+. .+.+. .+.+. .+.+.
<consen01> 1473 AC TGGGG TCCTTCC ATCTCTGGGGCCCC ACT CTCTT C TGTC TT
```

FIG. 17L

51 / 108

AA224590.RC	447	CCC-ATGGGAAGTGCC-CTGGN-ATCCC-TCTGCCC-TG
AA483522.RC	316	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
N20044.RC	273	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
2025350	267	CCC-ATGGGAAGTGCC-CTGGG
3212856	265	CCC-ATGGGAAGTGCC-CTGGG-A
1804959	250	CCC-ATG
AA244075	266	CCC-ATGGGAAGTGCC-CTGGG-ATCC--TCTGCC--TGTCC-TC--TG
1793273	223	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
1345563	220	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC
T40695.RC	214	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
R27969.RC	207	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
3144865	206	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
R72982	202	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
T86963.RC	165	CCC-ATGGGAANTGCC-CTG---ATCCC-TCTGCCC-TGTCC-TCC-TG
767739	148	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
647074	134	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TTC-TG
AA244018.RC	104	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
AA149993.RC	86	CCC-ATGGGAAGTGCC-ACNNGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
AA101562.RC	53	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
2223391	47	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
1447744	47	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
R01692.RC	46	CCCCATGGGAAGTGCCCACTGGG-ATCCCCCTGCCCCCTGCC-TCCCTG
R87078	29	GCT-GGAGTGCAGTGG-TATGAT-CTTGG-CTCACTG-TAACC-TCC-GC
AA101520.RC	29	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
T84017.RC	12	CCCNATGGGAAGTCCCCACTG---ATCCCCCTGCCCC-TGTCCCTCCCTG
1208791	13	CCC-ATGGGAAGTGCC-CTGGG-ATCCC-TCTGCCC-TGTCC-TCC-TG
1208826	13	CCC-ATGGGAAGTGCC-CTGGGTATCCC-TCTGCCC-TGTCC-TCC-TG
143613	12	CCC-ATGGGA-GTGCC-CTGGN-ATCC--TCTGCC--TGTCC-TCC-TG
241604	1	CTGCCC-TGTCC-TCC-TG
(SEQ ID NO: 185)		+. .+.+. .+.+. .+.+. .+.+. .+.+. .+.+. .+.+. .+.+. .+
<consen01>	1515	CCC ATGGGAAGTGCC ACTGGG ATCCC TCTGCCC TGTCC TCC TG
AA483522.RC	359	AATACAAGCTGACTGACATTGAA
N20044.RC	316	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
AA244075	306	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
1793273	266	AATACAAGCTGACTGACATTGACT
T40695.RC	257	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
R27969.RC	250	AATACAAGCTNACTNACATTGA
3144865	249	AATACAAGCTGACTGACATTGACTGTGTCTGTGG
R72982	245	AATACAAGCTGACTGACATTGACTGTGTCTGTGGGAAAATGGGAGCTCT
T86963.RC	206	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
767739	191	AATACAAGCTGACTGACATTGACTGTGTCTGTG
647074	177	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-GAAATGGG-AGCTTT
AA244018.RC	147	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
AA149993.RC	129	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
AA101562.RC	96	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
2223391	90	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
1447744	90	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
R01692.RC	94	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
R87078	72	CTCCCGGTTCAAGCCATTCTCCTGCCTCAGTCT-CCTGAGTA-GCTGGG
AA101520.RC	72	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
T84017.RC	58	AATANAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
1208791	56	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
1208826	57	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
143613	53	AATACAAGCTGACTGACAT-GACTGTGTCTGTGG-AAAATGGG-AGCTCT
241604	17	AATACAAGCTGACTGACATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
(SEQ ID NO: 186)	1	ATTGACTGTGTCTGTGG-AAAATGGG-AGCTCT
(SEQ ID NO: 187)	1	CCNCAGTCNCCTG-AGTAGCTG-GGATTG
(SEQ ID NO: 188)	1	GGG-AGCTCT
(SEQ ID NO: 189)	1	GGG-AGCTCT
<consen01>	1558	AATACAAGCTGACTGACATTGACTGTGTCTGTGG AAAATGGG AGCTCT

FIG. 17M

52 / 108

N20044.RC	364	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCGAAAAGGAT
AA244075	354	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
T40695.RC	305	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGA
R72982	295	TGTTGTGGGAGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGGT
T86963.RC	254	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
647074	225	TGTTGTGG-AGAGCATAGTAAANTTTTCAGAGG
AA244018.RC	195	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
AA149993.RC	177	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCGAAAAGGAT
AA101562.RC	144	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
2223391	138	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
1447744	138	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCGAAAAGGAT
R01692.RC	142	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
R87078	120	ATTGCGGG-TGCGTGCCACCATGCCTGGCTAATTTTGTGTTTTTGGTAG
AA101520.RC	120	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
T84017.RC	106	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
1208791	104	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
1208826	105	TGTTGTGG-A
143613	100	TGTTGTGG-AGAGCATAGTAA-TTTTCAGAGAACTTGAAGCCAAAAGGAT
241604	65	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGNT
816576	32	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
N54909.RC	28	CGGGTGCG-TGCCACCATGCCTGGCCTAATTTTGGGNTTTTAGTGGAGA
951273	10	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
2395956	10	TGTTGTGG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
(SEQ ID NO: 190) 608008	1	GG-AGAGCATAGTAAATTTTCAGAGAACTTGAAGCCAAAAGGAT
(SEQ ID NO: 191) AA196824.RC	1	GCCACCATGCCTGNCTGATTTTGGTGTTTTTAGTAGA
(SEQ ID NO: 192) T71041.RC	1	CCCCGGCCTAANTTTTGTGNTTTTAAGTAGAGACC
(SEQ ID NO: 193) 633873	1	AATTTTCAGAGAACTTNAAGCCAAAAGGGT
(SEQ ID NO: 194) 345566	1	AATTTTCAGAGAACTTGAAGCCAAAAGGAT
(SEQ ID NO: 195) R74032.RC	1	TNAGAAGAACAGGGTTTC
(SEQ ID NO: 196) H38626.RC	1	CACAGGGT
(SEQ ID NO: 197) 1578344	1	ACGTCAT
(SEQ ID NO: 198) W32430.RC	1	TG
		.....+ .+.....
<consen01>	1606	TGTTGTGG AGAGCATAGTAAATTTTCAGAGAACTTGAAGCGAAAAGGAT

FIG. 17N

53 / 108

```

N20044.RC      413 TT-AAAACCGCTGCTCT
AA244075      403 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
R72982        345 TTTAAAACCGCTGCTCTAAAGAAAAG-AAAAC--GGGAGGNTGGGCGC
T86963.RC     303 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
AA244018.RC   244 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
AA149993.RC   226 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
AA101562.RC   193 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
2223391       187 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
1447744       187 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
R01692.RC     191 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
R87078        169 AG-ACAGGGTTTACCATGTTGGTCG-GGCTGG--TC-TCAGACTCCTGA
AA101520.RC   169 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
T84017.RC     155 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGNCTGGGCGCA
1208791       153 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-A
143613        148 T--AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-A
241604        114 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-A
816576        81 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-A
N54909.RC     77 CC-AGGGTTCACCATGTTGGTCGGG-CTGGTC--TC-AAACTTCCTGAC
951273        59 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
2395956       59 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-A
608008        44 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-A
AA196824.RC   38 GA-CCAGGGTTTACCATGTTGGTCG-GGCTGG--TC-TCGGGCTCCTGA
T71041.RC     36 AG-GGTTTACCATGTTGGTCGGG-TCGGCT--CA-AANTCCCTGACC
633873        31 TT-AAAACCGCTGNTNTAAAGGAAAGGNAACCTTNGG-AGGNTTGGCGCA
345566        31 TT-AAAACCGCTGCTCTAAAGAAAAG-AAAAC--GG-AGGCTGGGCGCA
R74032.RC     19 CN-CCCCAGTTGGTCGGGCGCTGGTC-CCCCAG--AA-CTCCCGACCCCC
H38626.RC     9 TT-CCNCCAGNTTGGTCGGGCGCTGG-TCCTCA--GA-CTCCCTGACCCC
1578344       8 TG-TGTAAGGAAAAACAAACTAAGA-AAGTCC--AC-TCTCTGGTAAAA
W32430.RC     3 AG-GCAGGCGGATCACCTGAGGTCGG-GCTGGT--CT-CAGACTCCTGAC
R02367.RC     1 CN--CA-AANCCCCCTGCC
<consen01>    1655 TT AAAACCGCTGCTCTAAAGAAAAG AAAACT GG AGGCTGGGCGCA

AA244075      448 GTGGCTCAGCCTATAAT-CCCA-GA-GGCT-GAGGCA-GGCGG-ATC-A
R72982        392 AGTNGGCTTACGGCTAT-TAAT-CC-CCAG-AAGGTT-GAAGG-CAG-G
T86963.RC     348 AGTGCACGCCTGTAATCC-CAGA-GG-CTGA-GGCAGG-CGGAT-CAC-C
AA244018.RC   289 GTCGTCACGCCTATAATC-CCAG-AG-GCTG-AGGCAG-GCGGA-TCA-C
AA149993.RC   271 AGTGTGCGCCTGTAATC-CCAG-AG-GCTG-AGGCAG-GCGGA-TCA-C
AA101562.RC   238 AGTGCACGCCTATAATCC-CAGA-GG-CTGA-GGCAGG-CGGAT-CAC-C
2223391       232 GTGGCTCAGCCTATAA
1447744       232 G
R01692.RC     236 AGTGCACGCCTATAATCC-CAGA-GG-CTGA-GGCAGG-CGGAT-CAC-C
R87078        214 CCTCTTGATCCGCTGCC-TTGG-CC-TCCC-AAAGTG-ATGGGGATT-A
AA101520.RC   214 GTGCACGCCTATAATCCC-AGAG-GC-TGAG-GCAGGC-GGATC-ACC-T
T84017.RC     200 NGTGCACGCCTGTAATCC-CAGA-GG-CTGA-GGCAGG-CGGAT-CAC-C
N54909.RC     122 CTTCTTGATCCGCTGCC-TTGG-CCCTCCC-AAAGTG-ATGGG-ATT-A
951273        104 GTGGCTCANNNNNNNNNN-NNNN-NN-NNNN-NNNNNN-NNNN-NNN-N
AA196824.RC   83 CCTCTTGATCCGCTGCC-TTGG-CC-TCCC-AAAGTG-ATGGG-ATT-A
T71041.RC     81 CTCTTGANCCCGCTGCC-TTGG-CC-TCCCCAAAGTG-ATGGG-ATT-A
633873        79 ATGGCTNAAGGGCTGCC-TTGGNCC-TTCC-AAAGTGGATGGG-ATT-A
345566        76 GTGGCTCANGACTATAAT-CCCA-GA-GGCT-GAGGCA-GGCGG
R74032.RC     64 TGATCCCGCCCTGCCCTG-GCCC-TC-CCCC-AAAGTG-ATGGG-ATT-A
H38626.RC     54 CTTGGANCCCGCCTGCCCTTGG-CCCTCCCCAAAGTG-ATGGG-ATT-A
1578344       53 CATTTACCAAGCATATAA-ATTA-TG-AGGT-GCTGAT-TCATA-TGA-C
W32430.RC     48 CTCTTGATCCCGCTGCC-TTGG-CC-TCCC-AAAGTGNATGGG-ATT-A
R02367.RC     17 CNCTGATCCCGCTNCC-CTGG-CCNCCC-CAA-TG-ATGGG-ATT-A
R12602.RC     1 TCCCTGGCCCCCCC-CCCC-AA-GGAN-GGGATT-NCCCA-GAT-G
HUMGS02649    1 GATCCGCTGNN-TTGG-CC-TCCC-AAAGTG-ATGGG-ATT-A
R840069       1 T-A
<consen01>    1700 ATGGCTNAAGGGCTGCC TTGG CC TCCC AAAGTG ATGGG ATT A

```

FIG. 170

54 / 108

AA244075	491	C-C-TGAGGTC-AGG--A-GT-TAAGA-TCA-GC-CTGA-CCAGC-ATGG
R72982	435	C-C-GGNTCAC-CTG--A-AG-GTTCA-GGG-AT-TTCA-AGNTC-CAGC
T86963.RC	391	T-G-AGGTCAG-GAG--T-TC-AAGAT-CAG-CC-TGAC-CAACA-TGGA
AA244018.RC	332	C-T-GAGGTCA-GGA--G-TT-CAAGA-TCA-GC-CTGA-CCAAC-ATGG
AA149993.RC	314	C-T-GAGGTCG-GGA--G-TT-CGGGA-TCA-GC-CTGA-CCAAC-ATGG
AA101562.RC	281	T-G-AGGTCGG-GAG--T-TC-GGGAT-CAG-CC-TGAC-CAACA-TGGA
R01692.RC	279	T-G-AGGTCAG-GAG--T-TC-AGGAT-CAG-CC-TGAC-CAACA-TGGA
R87078	258	C-A-GATGTGA-GCC--A-CC-GTGCC-TAG-CC-AAGGGATGAGNATTT
AA101520.RC	257	G-A-GGTCGGG-AGT--T-CG-GGA
T84017.RC	243	T-G-AGGTCAG-GAG--T-TC-AAGAT-CAG-CC-TGAC-CAACA-TGGA
N54909.RC	166	CCA-GATGTGA-GCC--A-CC-GTGCC-TAG-CC-AAGG-ATGAG-ATTT
951273	147	N-N-NNNNNNN-NNN--N-NN-NNNNN-NNN-NN-NNNN-NNNNN-NNNN
AA196824.RC	126	C-A-GATGTGA-GCC--A-CC-GTGCC-TAG-CC-AAGG-ATGAG-ATTT
T71041.RC	125	C-A-GATGTGA-GCC--ANCC-GTGCC-TAG-CCCAAGG-ATGAG-ATTT
633873	124	C-AANATGTGA-GCC--A-NC-GTGCC-TAGNCC-AAGG-ATGAGGATTT
R74032.RC	108	CCA-GATGTGAAGCCC-A-CCCGTGCCCTAG-CCCAAGG-ATGAG-ANTT
H38626.RC	100	CCA-GATGTGA-GCCC-A-CCCGTGCCCTAG-CCCAAGG-ATGAG-ANTT
1578344	96	A-A-AAAGGAG-ATT--C-AC-TTTA-GTA-GC-TGCT-CTAAT-GCAT
W32430.RC	92	C-A-GATGTGA-GCC--A-CC-GTGCC-TAG-CCCAAGG-ATGAG-ATTT
R02367.RC	60	CCA-GATGTGA-NCCCCA-CCCGTCCC-CAG-CC-CAGG-ATGAG-ATTT
R12602.RC	40	T-G-AGCCCCC-NCC--C-CG-TCCCC-TAN-CC-CAAG-ATGAG-ATTT
HUMGS02649	38	C-A-GATGTGA-GCC--A-CC-GTGCC-TAG-CC-AAGG-ATGAG-ATTT
840069	3	C-A-GATGTGA-GCC--A-CC-GTGCC-TAG-CC-AAGG-ATGAG-ATTT
(SEQ ID NO: 203) 689191	1	C-A-GATGTGA-GCC--A-CC-GTGCC-TAG-CC-AAGG-ATGAG-ATTT
(SEQ ID NO: 204) 2300160	1	GTGCC-TAG-CC-AAGG-ATGAG-ATTT
(SEQ ID NO: 205) 2300168	1	GTGCC-TAG-CC-AAGG-ATGAG-ATTT
(SEQ ID NO: 206) 1320053	1	TGCC-TAG-CC-AAGG-ATGAG-ATTT
(SEQ ID NO: 207) 1669991	1	AAGG-ATGAG-ATTT
(SEQ ID NO: 208) 2728192	1	AAGG-ATGAG-ATTT
(SEQ ID NO: 209) 1274764	1	G-ATTT
(SEQ ID NO: 210) 1275979	1	G-ATTT
(SEQ ID NO: 211) 1271365	1	G-ATTT
(SEQ ID NO: 212) 1887285	1	ATTT
<consen01>	1743	C A GATGTGA GCC A CC GTGCC TAG CC AAGG ATGAG ATTT

FIG. 17P

55 / 108

AA244075	529	
R72982	473	TTGAC--CCCACATGGG-GGGA-AACCTT-ANTT-TT
T86963.RC	429	GAAAC--CCTACTGAAA-ATAC-AGAGTT-AGCC-AGGC-AT-GGT-GGT
AA244018.RC	370	AGAAA--CCCTACTAAA-AATA-CAAAGT-TAGC-CAGG-CA-TAG-TGG
AA149993.RC	352	AGAAA--CCCTACTGGG-AATA-CAGAGT-TGGC-CAGG-CA-TGG-TGG
AA101562.RC	319	GAAAC--CCTACTGGAA-ATAC-AAAGTT-AGCC-AGGC-AT-GGT-GGT
R01692.RC	317	GAAAC--CCTACTGGAA-ATAC-AAAGTT-AGCC-AGGC-AT-GGT-GGT
R87078	298	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGGGTTGGGAA-GAC-AGA
T84017.RC	281	GAAAC--CCTACTGAAA-ATAC-AGAGTT-AGCC-AGCA-TG-GTG-GTG
N54909.RC	205	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
951273	185	NNNN---TATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
AA196824.RC	164	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
T71041.RC	165	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
633873	165	TTAAA--GTATGTTTCA-NTTC-TGTGTN-ANGG-TTGG-AA-GAC-ANA
R74032.RC	152	TTAAA--GTATGTTTCA-GTTCCTGTGTCCATGG-TTGG-AA-GAC-AGA
H38626.RC	143	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ANGG-TTGG-AA-GAC-AGA
1578344	134	TCCAC--TTAAGTGAAT-ATTC-AAGGAT-TATT-TTGG-AA-GAC-AGA
W32430.RC	131	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
R02367.RC	102	TTAAAN-GTATGTTCCA-GTCC-TGTGTC-ATGG-TTGG-AA-GACCAGA
R12602.RC	78	TTAAAAAGTATGTTTCAAGTCC-T-TGTCCATGG-TGGG-AAANACCAGA
HUMGS02649	76	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
840069	41	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
689191	39	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
2300160	24	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-NA-GAC-AGA
2300168	24	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-NA-GAC-AGA
1320053	23	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
1669991	14	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
2728192	14	TTAAA--GTATGTTTCA-NTTC-TGTGTC-ATGG-TTNG-AA-GAC-AGA
1274764	6	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-NA-GAC-AGA
1275979	6	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
1271365	6	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
1887285	5	TTAAA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 213) 1862716	1	AA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 214) 3119215	1	AA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 215) 998106	1	AA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 216) 616405	1	AA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 217) 2453074	1	AA--GTATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 218) 2251286	1	TATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 219) 2451550	1	TATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 220) 1672494	1	TATGTTTCA-GTTC-TGTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 221) 2591955	1	TTTCA-GTTC-TGTGTC-ATGG-TTGG-NA-GAC-AGA
(SEQ ID NO: 222) 2259680	1	GTTC-TGTGTC-ATGG-TTGG-NA-GAC-AGA
(SEQ ID NO: 223) 1655649	1	C-GGCTCG-AGGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 224) 1734692	1	GTGTC-ATGG-TTGG-AA-GAC-AGA
(SEQ ID NO: 225) 786553	1	GG-TTGG-AA-GAC-AGA
(SEQ ID NO: 226) 1465664	1	GG-AA-GAC-AGA
(SEQ ID NO: 227) 2127319	1	A-GAC-AGA
(SEQ ID NO: 228) 1455536	1	AC-AGA
<consen01>	1781	TTAAA GTATGTTTCA GTTC TGTGTC ATGG TTGG AA GAC AGA

FIG. 17Q



56 / 108

T86963.RC	470	GCATG-CCTGTAATCC-CAGCTG-C-TCAGGAN---CCT-GGCAACA-AG
AA244018.RC	411	TGCAT-GCCTGTAATC-CCACCT-G-CTCTGT---TGC-CAGGCTC-CT
AA149993.RC	393	TGCAT-GCCTGTGGTC-CCAGCT-G-CTCAGGA---GCC-TGGCAAC-AA
AA101562.RC	360	GCATG-CCTGTAGTCC-CAGCTG-C-TCAGGAG---CCT-GGCAACA-AG
R01692.RC	358	GCATG-CCTGTAATCC-CAGCTG-C-TCAGGAG---CCT-GGCAACA-AG
R87078	341	GTAGGGAAGGTATGGGAAAAGG-TTCATGGGGGGGAAGGCAGAGGTTGA
T84017.RC	322	CATGC-CTGAATCCCA-GCTCTC-A-AGGANCC---TGG-CAACAAG-AG
N54909.RC	246	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
951273	224	GTAGG-AAGGATATGG-AAAA
AA196824.RC	205	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
T71041.RC	206	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
633873	206	GTAGG-NAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
R74032.RC	195	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
H38626.RC	184	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1578344	175	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGG
W32430.RC	172	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
R02367.RC	145	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
R12602.RC	124	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGN---AAG-CAGAGGT-GA
HUMGS02649	117	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
840069	82	GTAGG-AAGGNTATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
689191	80	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2300160	65	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2300168	65	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1320053	64	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1669991	55	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2728192	55	GTAGG-AAGGATATGG-AAAATG-T-CATNGGG---AAG-CAGAGGT-GA
1274764	47	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1275979	47	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1271365	47	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAGG-CAGAGGTNGA
1887285	46	NTAGG-AAGGATATGG-AAAAGG-T-CATGNGG---NNN-TCTGA
1862716	39	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
3119215	39	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
998106	39	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
616405	39	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2453074	39	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2251286	36	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2451550	36	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG
1672494	36	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2591955	32	GTAGG-AAGGATATNG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2259680	27	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1655649	24	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1734692	22	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
786553	15	GTAGG-AAGGATATGG-AAAAGG-N-CATGGGG---AAG-CAGAGGT-GA
1465664	11	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
2127319	8	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
1455536	6	GTAGG-AAGGATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
(SEQ ID NO: 229) 157587.RC	1	ATATGG-AAAAGG-T-CATGGGG---AAG-CAGAGGT-GA
(SEQ ID NO: 230) R86952.RC	1	G-T-CATGGGG---AAG-CAGAGGT-GA
(SEQ ID NO: 231) 1459939.RC	1	TGGGG---AAG-CAGAGGT-GA
<consen01>	1822	GTAGG AAGGATATGG AAAAGG T CATGGGG AAG CAGAGGT GA

FIG. 17R

57 / 108

T86963.RC	511 AG-C-AAAA-CT-CCAGC-TC-AAA
AA244018.RC	452 GA-G-CGTC-GA-GC
AA149993.RC	434 GA-G-CAAA-AC-TCCAG-CT-CAA-AA-AAAAA
AA101562.RC	401 AG-C-AAAA-CT-CCAGC-TC-AAA-AA-AAAAA-A
R01692.RC	399 AG-C-AAAA-CT-C
R87078	390 TTTC-ATGGGCT-CTGTGAA-TTTTGANGGTGA-AT-NG
T84017.RC	363 CA-A-AACT-CC-AGCTC-AA
N54909.RC	287 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
AA196824.RC	246 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
T71041.RC	247 NT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
633873	247 TT-N-ATNG-C
R74032.RC	236 NT-C-ATGG-CT-CTGTG-AA-NTT-GA-GGTGA-AT-GGTTCC-TTATT
H38626.RC	225 NT-C-ATGG-CN-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-NTANT
W32430.RC	213 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
R02367.RC	187 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGNTCC-TTATT
R12602.RC	165 TT-CCATGG-CT-CTGTG-AA-NTT-GA-GGTGA-AT-GGTTCC-TTATT
HUMGS02649	158 TT-C-ATGG-CT-CTGTG-AA-TTT-NA-GNTGA-AT-GGGTCC-TTATT
840069	123 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
689191	121 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
2300160	106 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
2300168	106 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1320053	105 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1669991	96 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
2728192	96 TT-C-ATGG-CT-CTGTN-AAATTT-NA-GGTGA-AT-GGTTCC-TTATT
1274764	88 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1275979	88 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1271365	90 TTTC-ATGGGCTTCTGTGAA-TTTTGA-GGTGA-ATTGGTTNC-CTTTA
1862716	80 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
3119215	80 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-T-ATT
998106	80 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
616405	80 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-G
2453074	80 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
2251286	77 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGANAT-GGTTCC-TTATT
1672494	77 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
2591955	73 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
2259680	68 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1655649	65 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1734692	63 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
786553	56 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1465664	52 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
2127319	49 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1455536	47 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
157587.RC	33 NT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
R86952.RC	22 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
1459939.RC	18 TT-C-ATGG-CT-CTGTG-AA-TTT-GA-GGTGA-AT-GGTTCC-TTATT
(SEQ ID NO: 232) 1433929	1 AT-GGTTCC-TTATT
(SEQ ID NO: 233) 1455495	1 CC-TTATT
<consen01>	1863 TT C ATGG CT CTGTG AA TTT GA GGTGA AT GGTTC TTATT

FIG. 17S

N54909.RC  
AA196824.RC  
T71041.RC  
R74032.RC  
H38626.RC  
W32430.RC  
R02367.RC  
R12602.RC  
HUMGS02649  
840069  
689191  
2300160  
2300168  
1320053  
1669991  
2728192  
1274764  
1275979  
1271365  
1862716  
3119215  
998106  
2453074  
2251286  
1672494  
2591955  
2259680  
1655649  
1734692  
786553  
1465664  
2127319  
1455536  
157587.RC  
R86952.RC  
1459939.RC  
1433929  
1455495  
(SEQ ID NO: 234) 878881  
(SEQ ID NO: 235) H49320.RC

<consen01>

[illegible]

FIG. 17T

<consen01>

```

373 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
332 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
333 AATTTACTTCTCTAGCTTANAATGGACCTTTTGAACTGG-AANANACCTT
322 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
311 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
299 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
273 AATTTACTTCTCTAGCTTACAATGGGCGCTTTTGAACTGG-NAAACACCTT
253 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAANACCTT
244 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
209 ANTTTACTTNNTAGCTTACAATGGACCTTTTGAACTGG-AAAACAACTT
207 AATTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACAA
192 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
192 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
191 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
173 AATT-ACCTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
166 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AA
159 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
154 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
151 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
149 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
144 AATTNACTTCNCTAGCNTACAATGGACCTNNNGAAGCTGGGAACANTCN
138 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
135 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
133 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
119 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
108 NATTTACNTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
104 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
61 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
55 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
33 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
34 AATTTACNTCTCTAGCTTACAATGGACCTTTTGAACTGG-AAAACACCTT
...++.+..+.+++.+.+++++.+++.....+++++.+++.+..+.
1949 AATTTACTTCTCTAGCTTACAATGGACCTTTTGAACTGG AAAACACCTT

```

FIG. 17U

N54909.RC  
AA196824.RC  
T71041.RC  
R74032.RC  
H38626.RC  
W32430.RC  
R02367.RC  
R12602.RC  
HUMGS02649  
840069  
2300160  
2300168  
1320053  
1275979  
2591955  
2259680  
1655649  
1734692  
786553  
1465664  
2127319  
1455536  
157587.RC  
R86952.RC  
1459939.RC  
1433929  
1455495  
878881  
H49320.RC

```

422 GTCTGCATTCACTTTAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
381 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
382 GTCTGCATTCACTTTAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
371 GTCTGCATTCACTTTAAATGTCAAAC-TAATTTTT-A
360 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAANTT
348 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
322 GTCTGCATTCACTTTAAAATGTCAAACCTAATTTTTTATAATAAATGTT
302 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
293 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTN-ATAATAANTGTT
258 GTCTGC
241 GTCTGCATTCACTT
241 GTCTGCATT
240 GTCTGCATTCACTT
221 GTCTGCATTCACTTTAAAT
208 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAAT
203 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAAT
200 GTCTGCATTCACTTTAAATGTAAAATT
198 GTCTGCATTCAAC
194 GTCNGCATNCACNTAAAANGNCAAAC-NAATTNNN-ATAANAATGNT
187 GTCTGCATTCACTTTG
184 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
182 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
168 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
157 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
153 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
110 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
104 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
82 GTCTGCATTCACTTTAAAATGTCAAAC-TAATTTTT-ATAATAAATGTT
83 GTCTGCATTCACTTTAAAATGTT
+++ .++++.+++++.+++++...++ ..+.++ ++.++++.+
1998 GTCTGCATTCACTTTAAAATGTCAAAC TAATTTTT ATAATAAATGTT

```

N54909.RC  
AA136824.RC  
T71041.RC  
H38626.RC  
W32430.RC  
R02367.RC  
R12602.RC  
HUMGS02649  
786553  
2127319  
1455536  
157587.RC  
R86952.RC  
1459939.RC  
1433929  
1455495  
878881

```

470 TATTTTCACATTG
429 TATTTTCACATTGA
430 TATTTTCACATCCAACCAAAA
408 TATTTTCAC
396 TATTTTCACATTG
372 TATTTTCACAAAAAAA
350 TATTTTCACATTGAAAAAAA
341 TATTTTCACATTGGAAA
242 NANNINCACANTGAAAA
232 TATTTTCACATTG
230 TATTTTCACATTGAAAAAA
216 TATTTTC
205 TATTTTCAC
201 TATTTTCACATTG
158 TATTTTCACATTG
152 TATTTTCACATTG
130 TATTTTCACATTG
    .+.+.+++++.+.+++++
2046 TATTTTCACATTGAAAAAAA

```

**SUBSTITUTE SHEET (RULE 26)**

61 / 108

OLI2162 (35936.f1)

SEQ ID NO:78

TCGCGGAGCTGTGTTCTGTTTCCC

OLI2163 (35936.p1)

SEQ ID NO:79

TGATCGCGATGGGGACAAAGGCGCAAGCTCGAGAGGAACTGTTGTGCCT

OLI2164 (35936.f2)

SEQ ID NO:80

ACACCTGGTTCAAAGATGGG

OLI2165 (35936.r1)

SEQ ID NO:81

TAGGAAGAGTTGCTGAAGGCACGG

OLI2166 (35936.f3)

SEQ ID NO:82

TTGCCTTACTCAGGTGCTAC

OLI2167 (35936.r2)

SEQ ID NO:83

ACTCAGCAGTGGTAGGAAAG

**FIG. 18**

**A33 human - A33 antigen precursor - Homo sapiens (319 aa)**

**Score = 246 (86.6 bits), Expect = 2.8e-19, P = 2.8e-19**

**Identities = 81/268 (30%), Positives = 131/268 (48%), at 121,17, Frame = +1**

**SUBSTITUTE SHEET (RULE 26)**

DNAA040628 121 LALGSVTVHSSEPEVRIPENNPKLSAYSGFSPR---VEW-KFDQGDTRRLVC--YNN  
 SEQ ID NO:84

17 VTVDAISVETPQDVLRAQGSVTLPTCYHTSTSSREGLIQWDKLLLTHTERVVIVWPFNS  
A33 human  
SEQ ID NO:85

DNA40628 283 K--ITAS-YEDRVTFI-----PTGITFKSVTREDTGTYTCMVS---EEGGSNSYGEVVKVK

DNA40628 427 LVLVPPSKPTVNIPSSATIGNRAVLTCSEQDGSPPSEYTFWKDGI VMP TNP KSTRAF SN  
 \* . \*\*\*\*\* \* \*\*\* \*\*\* .. \*\*\*\* . \*. \* . \*

A33 human 135 LLVLVPPSKPEGIETII GNN IQL TCQSKEGSP TP Q YSWKRYN IL NQE P-----

DNA40628 607 SSYVLNPTTGELV-FDPLSASDTGEYSCEARNGYGTPMTSNAVRMEAVERNVGV---IVA  
A33 human 187 ---LAQPASGQPSVLKNISTDTSGYYICTSSNEEGTQFCNITVAVRSPSMNVALYVGIIV

DNA40628 775 AVLVTLLGILVFGIWFAYSRGHFDRT--KKGTSSKKVIYSQP

A33 human 244 GVVAALIIIGIIY--CCCCRGKDDNTEDKEDARPNREAYEEP

**FIG. 19A**

63/108

Score = 245 (86.2 bits), Expect = 3.6e-19  
 Identities = 83/273 (30%), Positives = 131/273 (47%), at 112,12, Frame = +1

```

DNA40628 112 LCSSL--ALGSVTVHSSEPEVRIPENNPVKLSAYSGFSSPR---VEW-KFDQGDTTRLVC
SEQ ID NO:86
A33_human 12 LCAVRVTVDAISVETPQDVLRASQKSVTLPCITYHTSTSSREGLIQWDKLLLTHTERVVI
SEQ ID NO:87

DNA40628 274 --YNNK--ITAS-YEDRVTFI-----PTGITFKSVTREDTGTYTCMVSEEGGNSYGEVK
A33_human 72 WPFNSKNYIHGELYKNRVSISNNAEQSDASITIDQLTMADNGTYECSVSLMS-DLEGNTK

DNA40628 421 --VKLIIVLPSPKPTVNIPSSATIGNRAVLTCSEQDGSPPSEYTWFKDGI VMP TNP KSTR
A33_human 131 SRVRLIVLPSPKPEGIEGETIIGNNIQLTCQSKESPTPQYSWKRYNILNQEPP---

DNA40628 595 AFSNSSYVLNPTTGELV-FDPLSASDTGEYSCEARNGYGTPMTSNAVRMEAVERNVGV--
A33_human 187 -----LAQPASGQPVSLKNISTDTSGYIYICTSSNEEGTQFCNITVAVRSPSMNVALYV

DNA40628 766 -IVAAVLVTLILLGILVFGIWFAYSRGHFDRT--KKGTSKKKVIYSQP
A33_human 240 GIAVGWVAALIIIGIIY---CCCCRGKDDNTEDKEDARPNREAYEEP

```

FIG. 19B



64 / 108

```

>< /usr/seqdb2/sst/DNA/Dnaseqs.full/ss.DNA37150 (2943 bases)
< good sequence: 1-2943 (2943 bases)
< insert: 91-2845 (2755 bases), 10 regions found
< 5' PRK5 + lk: 36-911, 40 matches (100%), 40 consec
< 5' PRK5 + lk: 17-901, 21 matches (87%), 14 consec, 1 gap
< 3' PRK5 + lk: 2846-1566, 66 matches (100%), 66 consec
< PRK5D: 2848-1366, 81 matches (100%), 81 consec
< PRK5D: 1-885, 78 matches (100%), 78 consec
< PRK5B: 2848-1357, 81 matches (100%), 81 consec
< PRK5B: 1-885, 37 matches (92%), 30 consec, 1 gap
< PRK5B: 36-911, 28 matches (100%), 28 consec
< 3' cDNA linker: 2846-1566, 26 matches (100%), 26 consec
< 5' cDNA linker: 75-950, 16 matches (100%), 16 consec

```

SEQ ID NO:236

```

GGGGTTAGGGAGGAAGGAATCCACCCCAACCCCTTTCTCTCCTTCCCT
GGCTTCGGACATTGGAGCACTAATGAATGAACTTGAATGTGTGTGGCGAGCAGGATGGTC
GCTGTTACTTTGTGATGAGATCGGGGATGAATGCTCGCTTTAAAA
<MET {trans=1-s, dir=f, res=1}>

```

SEQ ID NO:237

```

ATGCTGCTTTGGATTCTGTGCTGGAGACGTCTCTTTGTTTGGCGTGGAAACGTTACA
GGGACGTTTGCAAGAGAGAGATCTGTTCTGCAATGAGATAGAAAGGGACCTACACGTA
GACTGTGAAAAAAGGGCTTCACAAGTCTGCAGCGTTTCACTGCCCGACTTCCAGTTT
TACCATTATTTCTGCATGGCAATCCCTCACTCGACTTTTCCCTAATGAGTTCGCTAAC
TTTTATAATGCGGTTAGTTTGCACATGGAAACAATGGCTGCATGAATCGTTCGGGG
GCTTTCTGGGCTGCAGCTGGTGAAAGGCTGCACATCAACAACAAGATCAAGTCT
TTTCGAAAGCAGACTTTCTGGGCTGGACGATCTGGAATATCTCCAGGCTGATTTAAT
TTATTACGAGATATAGACCGGGGGCTTCCAGGACTTGAACAAGCTGGAGGTGCTCAT
TTAAATGACAAATCTCATCAGCACCTACCTGCCAACGTGTTCCAGTATGTGCCCATCACC
CACCTCGACCTCCGGGGTAACAGGCTGAAACGCTGCCCTATGAGGAGGTCTTGGAGCAA
ATCCCTGGTATTGGGGAGATCCTGCTAGAGGATAACCCCTGGGACTGCACCTGTGATCTG
CTCTCCCTGAAAGAATGGCTGGAACAATTCCTCAAGAATGCCCTGATCGGCCGAGTGGTC
TGCGAAGCCCCCACCAGACTGCAGGGTAAAGACCTCAATGAACCAACCGAAGAGACTTG
TGTCCTTTGAAAAACCGAGTGGATTCTAGTCTCCCGGGCCCCCTGCCCAAGAGAGACC

```

FIG. 20A

65 / 108

TTTGTCTCCTGGACCCCTGCCAACTCCTTTCAAGACAAATGGGCAAGAGGATCATGCCACA  
CCAGGGTCTGCTCCAAACGGAGGTACAAAGATCCCAGGCAACTGGCAGATCAAAATCAGA  
CCCACAGCAGCGATAGCGACGGGTAGCTCCAGGAACAACCCCTTAGCTAACAGTTTACCC  
TGCCCTGGGGCTGCAGCTGCGACCCACATCCCAGGGTCGGGTTTAAAGATGAACCTGCAAC  
AACAGGAACGTGAGCAGCTTGGCTGATTGAAGCCCAAGCTCTCTAACGTGCAGGAGCTT  
TTCTACGAGATAACAAGATCCACAGCATCCGAAATCGCACTTTGTGGATTACAAGAAC  
CTCATTTCTGTTGGATCTGGCAACAATAACATCGCTACTGTAGAGAAACAACACTTTCAAG  
AACCTTTTGGACCTCAGGTGGCTATACATGGATAGCAATTACCTGGACACGCTGTCCCGG  
GAGAAATTGCGGGGCTGCAAAACCTAGATACCTGAACGTGGAGTACAACGCTATCCAG  
CTCATCTCCCGGCACTTTCAATGCCATGCCCCAACTGAGGATCCTCATTTCTCAACAAC  
AACCTGCTGAGGTCCCTGCCCTGTGGACGTGTTCCGTGGGTCTCGCTCTCTAAACTCAGC  
CTGCACAACAATTACTTTCATGTACCTCCCGTGGCAGGGGTGCTGGACCACTTAACCTCC  
ATCATCCAGATAGACCTCCACGGAAACCCCTGGGAGTGCTCTGCACAAATTGTGCCTTTC  
AAGCAGTGGGCAGAACGCTTGGGTTCCGAAGTGCTGATGAGCGACCTCAAGTGTGAGACG  
CCGTGAACCTTTAGAAAGGATTTTCATGCTCCTCTCCAATGACGAGATCTGCCCTCAG  
CTGTACGCTAGGATCTCGCCACGTTAACTTCGCACAGTAAAAACAGCACTGGGTTGGCG  
GAGACCGGACGCACTCCAACCTCCTACCTAGACACCAAGGGTGTCCCATCTCGGTGTG  
GTCCCGGACTGCTGCTGGTGTGTCACCTCCGCTTACCGTGGTGGGCATGCTCGTG  
TTTATCTGAGGAACCGAAGCGGTCCAAGACGAGATGCCAACTCCTCCGCGTCCGAG  
ATTAAATCCCTACAGACAGTCTGTGACTCTTCTACTGGCACAATGGGCCTTACAACGCA  
GATGGGCCCCACAGAGTGATGACTGTGGCTCTCACTCGCTCTCAGACTAAGACCCCAAC  
CCCAATAGGGAGGCGCAGAGGGAAGCGGATACATCCTTCCCCACCGCAGGCAACCCCGGG  
GCTGGAGGGCGGTACCCAAATCCCCCGCCCATCAGCCTGGATGGGCATAAGTAGATAA  
ATAACTGTGAGCTCGCACAAACCGAAAGGGCCCTGACCCCTTACTTAGCTCCCTCCTTGAA  
CAAAGAGCAGACTGTGGAGAGCTGGGAGAGCGCAGCCAGCTCGCTCTTTGCTGAGAGCCC  
CTTTTGACAGAAAGCCACGACGCCCTGCTGGAAGAACTGACAGTGCCCTCGCCCTCGG  
CCCCGGGCTGTGGGTGGATGCGCGGTTCTATACATATATACATATATCCACATCT  
ATATAGAGAGATAGATATCTATTTTCCCTGTGGATTAGCCCCGTGATGGCTCCCTGTT  
GGCTACGCAGGGATGGGCAGTTGCACGAAGGCATGAATGTATTGTAATAAGTAACCTTGG  
ACTTCTGAC

SEQ ID NO:237

FIG. 20B

></usr/seqdb2/sst/DNA/DNAseqs.min/ss.DNA37150  
><subunit 1 of 1, 696 aa, 0 stop  
><MW: 77735, pI: 6.36, NX(S/T): 6

SEQ ID NO: 238

SEQ ID NO:238 MLLWILLLETSLCEAAGNVGTGVCCKEICSCNEIEGDLHVDCEKKGFTSLQRFAPTSTQ  
FYHLFLHGNSLTRLPNEFANFYNAVSLHMENNGLHEIVPGAFLGLQLVKRLHINNKI  
KSFRKQTFGLDDLEYLQADENLLRDIIDPGAQDLNKLEVLILNDNLISTLPANVFQYV  
PITHLDLRGNRLKTLPEEEVLEQIPGIAEILLEDPWDCTCDLLSLKEWLENIPKNALI  
GRVVEAPTRLQGKDLNETTEQDLCPKKNRVDSSLPAPPAQEETFAPGPLPTPFKTNGQ  
EDHATPGSAPNGGTKIPGNWQIKIRPTAAIATGSSRNKPLANSIPCPGGCSDHIPGSG  
LKMNCNRRNVSSLADLKPKLSNVQELFLRDNKIHSIRKSHFVDYKNLILLDLGNNNIAT  
VENNTFKNLLDLRWLYMDSNYLDTLSREKFAGLQNLLEYLNVEYNAIQILPGTFNAMPK  
LRILILNNLLRSLPVDVFAGVSLSKLSLHNNYFMYLPVAGVLDQLTSIIQIDLHGNPW  
ECSTIVPFKQWAERLGSEVMSDLKCETPVNFFRKDFMLLSNDEICPOLYARISPTLT  
SHSKNSTGLAETGTHSNSYLDTSRVSISVLVPGLLLVFVTSFTVVGMLVFILNRKRS  
KRRDANSSASEINSLQTVCDSSYWHNGPYNADGAHRVYDCGSHSLSD

FIG. 21

DNA37150 770 PGIAEILLEDNPWDCTCDLLSLKWELENI PKNALIGRVVCEAPTRLQCKDLNETTEQ  
 . \* \*\* \* \* \* \*  
 SLIT DROME 442 KSMKTVHLAKNPFICDCNLRLWLADYLHKNPJETSGAR--CESPKRMHRRRIESLREE  
 . . . . .

FIG. 22A

```

Score = 177 (62.3 bits), Expect = 4.2e-07, Sum P(2) = 4.2e-07
Identities = 44/127 (34%), Positives = 67/127 (52%), at 1229,105, Frame = +2

DNA37150 1229 LKMNCNNRNVSSLADLKPKLSNVQELFLRDNKHISRKSHFVDYKNLILLDLGNNNIATV
      *.. ** * * * .. * * * .. * * * .. * * * .. * * * .. * * * .. * * *
SEQ ID NO:241 SLIT_DROME 105 LELQGNLTVIYETDFQ-RLTKLRMLQLTDNQIHTIERNFSQDLVSLERLDISNNVITTV

DNA37150 1409 ENNTEFKNLLDLRLWLYMDSNYDLTLSREKFAGLQNLEYLNVEYNAIQILPLPGTFNAMPKLR
      ** ** * * * .. * * * .. * * * .. * * * .. * * * .. * * * .. * * *
SLIT_DROME 164 GRRVFKGAQSLRSLQLDNNQITCLDEHAFKGLVEILEILTNNNNILTSLPHNIFGGIGRLR

DNA37150 1589 ILILNHN
      * * * .. *
SLIT_DROME 224 ALRLSDN

Score = 160 (56.3 bits), Expect = 2.5e-16, Sum P(4) = 2.5e-16
Identities = 48/146 (32%), Positives = 66/146 (45%), at 251,299, Frame = +2

DNA37150 251 CSCNEIEGLHVDCEKKGFTSLQRFAPTQSQFYHLFLHGNLSLTRLFPNEFANFYNAVSLH
      * * * * * .. * * * .. * * * .. * * * .. * * * .. * * * .. * * *
SEQ ID NO:242 SLIT_DROME 299 CRC-----ADGIVDCREKSLTSVP-VTLPPDDTT-DVRLEQNQFITELEPPKSFSSFRRLRRID

DNA37150 431 MENNGLHEIVPGAFLGLQLVKRLHNNNKKIKSFRKQTFGLGLDDLEYLQADFNLLRIDPG
      . * * * * * .. * * * .. * * * .. * * * .. * * * .. * * * .. * * *
SLIT_DROME 353 LSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLP SGVFKGLGSLRLLLLNANEISCIRKD

DNA37150 611 AFQDLNKLVLILNDNLISLTPANVF
      **..**.. * * * .. * * * .. * * * .. * * * .. * * * .. * * * .. * * *
SLIT_DROME 413 AFRDLHLSLLSLYDNNIQSLANGTF

Score = 156 (54.9 bits), Expect = 3.2e-18, Sum P(3) = 3.2e-18
Identities = 45/146 (30%), Positives = 72/146 (49%), at 1448,747, Frame = +2

DNA37150 1448 LYMDSNYDLTLSREKFAGLQNLEYLNVEYNAIQILPLPGTFNAMPKLRILILNNNLRLSLP
      **..**.. * * * .. * * * .. * * * .. * * * .. * * * .. * * * .. * * *
SLIT_DROME 747 LYLESNEIEQIHVERIRHRLSLTRLDLSNQITILSNYTFANLTSLTLIISYNKLQCLQ

```

**FIG. 22B**



70 / 108

Score = 46 (16.2 bits), Expect = 2.5e-16, Sum P(4) = 2.5e-16  
 Identities = 13/33 (39%), Positives = 17/33 (51%), at 704,528, Frame = +2

DNA37150 704 THLDLRGNRLKTLPYEEVLEQIPGIAEILLEDN  
 \* . \* \* \* \* . \* \* \* \* \*  
 SEQ ID NO:246 SLIT\_DROME 528 TTVDCTGRRRLKEIPRDIPLHT-----TELLLNDN

Score = 40 (14.1 bits), Expect = 3.2e-18, Sum P(3) = 3.2e-18  
 Identities = 8/19 (42%), Positives = 11/19 (57%), at 2504,1347, Frame = +2

DNA37150 2504 PLLTESPARPCWKN\*QCPR  
 \* . \* \* \* \* . \* \* \* \* \*  
 SEQ ID NO:247 SLIT\_DROME 1347 PHIKEFPVDPCLN-KCRR

FIG. 22D





Score = 173 (60.9 bits), Expect = 4.4e-17, Sum P(3) = 4.4e-17  
Identities = 43/176 (24%), Positives = 85/176 (48%), at 413,323, Frame = +2

DNA37150 413 NAVSLHMENGLHEIVPGAFLGLQVVKRLHINNKKIKFRKQTFGLGLDDLEYLQADFNLL  
SEQ ID NO:250 P\_R25079 323 DTTDVRLEQNFITELPPKSFSSFRRLRRRIDLSNNNISRIAHDAISGLKQLTTIVLYGNKI  
DNA37150 593 RDIDPGAQDLNKLKLEVLILNDNLISLTPANVFQYV-PITHLDLRGNRLKTLPEYEEVLEQI  
P\_R25079 383 KDLPSGVFKGLSLRLILLNANEISCIKDAFRDLHSLSLSLYDNNIQSVA-NGTFDAM  
DNA37150 770 PGIAEILLEDNPNWDCDCLLSLKEWLENIPKNALIGRVVCEAPTRLQKDLNETTEQ  
P\_R25079 442 KSMKTVHLAKNPFIDCNLRWADYLHKNPIETSGAR--CESPKRMHRRRIESVREE

Score = 157 (55.3 bits), Expect = 2.1e-15, Sum P(4) = 2.1e-15  
Identities = 47/146 (32%), Positives = 66/146 (45%), at 251,299, Frame = +2

DNA37150 251 CSCNEIEGDLHVDCEKKGFTSLQRFAPTQFYHLFLHGNSLTRLPNEFANFYNAVSLH  
SEQ ID NO:251 P\_R25079 299 CRC-----ADGIVDCREKSLTSVP-VTLPPDDTT-DVRLEQNFITELPPKSFSSFRRLRRID  
DNA37150 431 MENNGLHEIVPGAFLGLQVVKRLHINNKKIKFRKQTFGLGLDDLEYLQADFNLLRDIDPG  
P\_R25079 353 LSNNNISRIAHDAISGLKQLTTIVLYGNKIKDLP SGVFKGLGSLRLILLNANEISCIRKD  
DNA37150 611 AFQDLNKLKLEVLILNDNLISLTPANVF  
P\_R25079 413 AFRDLHSLSLSLYDNNIQSVANGTF

Score = 150 (52.8 bits), Expect = 4.4e-17, Sum P(3) = 4.4e-17  
Identities = 45/146 (30%), Positives = 72/146 (49%), at 1448,747, Frame = +2

DNA37150 1448 LYMDSNYLDTLSREKFAGLQNLVEYNAIQLILPGTFNAMPKLRILILNNNLLRSLP  
SEQ ID NO:252 P\_R25079 747 VYLESNEIEQIHRYERIRHLRSLRDLNNSNQITILSNYTFANLTKLSRLIISYNKLQCLQ

FIG. 23B

73 / 108

DNA37150 1628 VDVFAGVS-LSKLSLHNNYFMYLPVAGVLDQLTSTIIQIDLHGNPWCSCCTIVPFKQWAER  
 P\_R25079 807 RHALSGLNLRVVSLSHGRIISMLPEASFED-LKSLTHIALGNSPLYCDCGLKWFSDWIKL  
 DNA37150 1805 LGSEVLMSDLKCETPVNFFRKDFMLLS  
 P\_R25079 866 --DYVEPGIARCAEPEQM--KDKLILS

Score = 117 (41.2 bits), Expect = 0.75, Sum P(2) = 0.53  
 Identities = 40/164 (24%), Positives = 72/164 (43%), at 494,746, Frame = +2

DNA37150 494 RLHINNKKIKSFRKQTFGLGLDDLEYLQADFNLRLRIDPGAFQDLNKLEVLILNDNLISLTL  
 SEQ ID NO:253 P\_R25079 746 QVYLESNEIEQIHRYIRHLRSLTRLDLSNNQITILSNYTFANLTKLSRLIISYNKLQCL  
 DNA37150 674 PANVFQVVP-ITHLDLRGNRLKTLPEYEVLEQIPGIAELLLEDNPDWDCDCLLSLKEWLE  
 P\_R25079 806 QRHALSGLNNLRVVSLSHGRIISMLP-EASFEDLSLTHIALGNSPLYCDCGLKWFSDWIK  
 DNA37150 851 NIPKNALIGRVVCEAPTRLQCKDLNETTEQD-LCPLKNRVDSSLPA  
 P\_R25079 865 LDYVEPGIAR--CAEPEQMKDKLILSTPSSFVC--RGRVRNDILA

Score = 87 (30.6 bits), Expect = 1.2e-10, Sum P(3) = 1.2e-10  
 Identities = 28/103 (27%), Positives = 46/103 (44%), at 1229,551, Frame = +2

DNA37150 1229 LKMNCNRRNVSSSLADLKPKLSNVQELFLRDNKKIHSIRKSHFVDYKNLILLDLGNNNIATV  
 SEQ ID NO:254 P\_R25079 551 LLLNDNELGRISDGLFGRPLHLVKLELKRNLQLTGIEPNAFEGASHIQELQNGENKIKEI  
 DNA37150 1409 ENNTFKNLLDLRWLYMDSNYLDLTSREKFAQLQNLEYLNVEYN  
 P\_R25079 611 SNKMFLGLHLQKLTNLNDNQISCVMPGSFEHLNLSLTLNLSN

FIG. 23C

74 / 108

Score = 46 (16.2 bits), Expect = 2.1e-15, Sum P(4) = 2.1e-15  
Identities = 13/33 (39%), Positives = 17/33 (51%), at 704,528, Frame = +2

SEQ ID NO:255	P_R25079	DNA37150	704	528
	TTTVDCTGRRLLKEIPRDIPLHT----	THLDLRGNRLKTLPYEEVLEQIPGIAEILL	THLDLRGNRLKTLPYEEVLEQIPGIAEILL	EDN

Score = 40 (14.1 bits), Expect = 4.4e-17, Sum P(3) = 4.4e-17  
Identities = 8/19 (42%), Positives = 11/19 (57%), at 2504,1347, Frame = +2

DNA37150 2504 PLLTESPARPCWKN\*QCPR  
 \* . \* \* \* \*  
 SEQ ID NO:256 1347 PHIKEEVPDPLEN-KCRR  
 \* . \* \* \* \*

FIG. 23D

75/108

genseq working...  
>T73996 yc81b12.r1 Homo sapiens cDNA clone 22385 5' similar to SP:B36665  
<379 bases SEQ ID NO:257  
GCACCTTTGTGGATTACAAGAACCTCATTTCTGTTGGATCTGGGCAACAATAACATCGCTA  
CTTGAGAGAACAAACACTTCAAGAACCTTTTGACCTCAGGTGGCTATACATGGATAGC  
AATTACCTGGACACGCTGTCCCGGAGAAATTCCGGGGCTGCAAAACCTAGAGTACCTG  
AACGTGGAGTACAACGCTATCCAGCTCATCTCCCGGGCACTTCAATGCCATGCCCCAA  
CTGAGGATCCTCATTTCTCAACAACAACCTGCTGAGGTCCCTGCCGTGGGACGTGTTGCG  
TGGGGTCTCGCTCTCTTAAACTCAGCCTGCACAACAATTACTTCATGTACCTCCCGGTGG  
GCAGGGGGTGCTGGGACC

stdin: END

FIG. 24

SEQ ID NO:261

76 / 108

GTAAGTGAAGTCAGGCTTTTCATTTGGGAAGCCCCCTCAACAGAATTCGGTCATTCTCCA  
AGTT

SEQ ID NO:262}

ATGGTGGACGTACTTCTGTTGTTCTCCCTCTGCTTGCTTTTTTACATTAGCAGACCGGAC  
TTAAGTCACAACAGATTATCTTTCATCAAGGCAAGTTCCATGAGCCACCTTCAAAGCCTT  
CGAGAAGTGAACTGAACAACAATGAATTGGAGACCATTCCAAATCTGGGACCAGTCTCG  
GCAAATATTACACTTCTCTCCTTGGCTGGAAACAGGATTGTTGAAATACTCCCTGAACAT  
CTGAAAGAGTTTCAGTCCCTTGAACCTTTGGACCTTAGCAGCAACAATATTTAGAGCTC  
CAAAGTGCATTTCCAGCCCTACAGCTCAAATATCTGTATCTCAACAGCAACCGAGTCACA  
TCAATGGAACCTGGGTATTTTGACAATTTGGCCAACACACTCCTTGTGTTAAAGCTGAAC  
AGGAACCGAATCTCAGCTATCCACCCAAGATGTTTAAACTGCCCCAACTGCAACATCTC  
GAATTGAACCGAAACAAGATTAAAAATGTAGATGGACTGACATTCCAAGGCCTTGGTGCT  
CTGAAGTCTCTGAAAATGCAAAGAAATGGAGTAACGAAACTTATGGATGGAGCTTTTTGG  
GGCTGAGCACTGAAAATTTTGACAGCTGGACCATAACAACCTAACAGAGATTACCAA  
GGCTGGCTTTACGGCTTGCTGATGCTGCAGGAACCTCATCTCAGCCAAAATGCCATCAAC  
AGGATCAGCCCTGATGCCTGGGAGTTCTGCCAGAAGCTCAGTGAGCTGGACCTAACTTTC  
AATCACTTATCAAGGTTAGATGATTCAAGCTTCCTTGGCCTAAGCTTACTAAATACACTG  
CACATTGGGAACAACAGAGTCAGCTACATTGCTGATTGTGCCTTCCGGGGGCTTTCCAGT  
TTAAAGACTTTGGATCTGAAGAACAATGAAATTTCTGGACTATTGAAGACATGAATGGT  
GCTTTCTCTGGGCTTGACAACTGAGGCGACTGATACTCCAAGGAAATCGGATCCGTTCT  
ATTACTAAAAAAGCCTTCACTGGTTTGGATGCATTGGAGCATCTAGACCTGAGTGACAAC  
GCAATCATGTCTTTACAAGGCAATGCATTTTCACAAATGAAGAACTGCAACAATTGCAT  
TTAAATACATCAAGCCTTTTGTGCGATTGCCAGCTAAAATGGCTCCCACAGTGGGTGGCG  
GAAAACAACCTTTCAGAGCTTTGTAAATGCCAGTTGTGCCCATCCTCAGCTGCTAAAAGGA  
AGAAGCATTTTGTGTTAGCCAGATGGCTTTGTGTGTGATGATTTTCCCAAACCCAG  
ATCAGGTTTCAGCCAGAAACACAGTCGGCAATAAAAGGTTCCAATTTGAGTTTCATCTGC  
TCAGCTGCCAGCAGCAGTGATTCCCAATGACTTTTGCTTGAAAAAAGACAATGAACATA  
CTGCATGATGCTGAAATGGAAATTTATGCACACCTCCGGGCCCCAAGGTGGCGAGGTGATG  
GAGTATACCACCATCCTTCGGCTGCGCGAGGTGGAATTTGCCAGTGAGGGGAAATATCAG  
TGTGTCATCTCCAATCACTTTGGTTTATCTCTGTCAAAGCCAAGCTTACAGTAAAT  
ATGCTTCCCTCATTACCAAGACCCCCATGGATCTCACCATCCGAGCTGGGGCCATGGCA  
CGCTTGGAGTGTGCTGCTGTGGGGCACCCAGCCCCCAGATAGCCTGGCAGAAGGATGGG  
GGCACAGACTTCCAGCTGCACGGGAGAGACGCATGCATGTGATGCCCGAGGATGACGTG  
TTCTTTTATCGTGGATGTGAAGATAGAGGACATTGGGGTATACAGCTGCACAGCTCAGAAC  
AGTGCAGGAAGTATTTAGCAATGCAACTCTGACTGTCCTAGAAACACCATCATTTTTG

CGGCCACTGTTGGACCGAACTGTAACCAAGGGAGAAACAGCCGTCTACAGTGCATTGCT  
GGAGGAAGCCCTCCCCCTAAACTGAACTGGACCAAAGATGATAGCCCATTGGTGGTAACC  
GAGAGGCACTTTTTTGCAGCAGGCAATCAGCTTCTGATTATTGTGGACTCAGATGTCAGT  
GATGCTGGGAAATACACATGTGAGATGTCTAACACCCTTGGCACTGAGAGAGGAAACGTG  
CGCCTCAGTGTGATCCCCACTCCAACCTGCGACTCCCCTCAGATGACAGCCCCATCGTTA  
GACGATGACGGATGGGCCACTGTGGGTGTCGTGATCATAGCCGTGGTTTGTGTGTGGTG  
GGCACGTCACTCGTGTGGGTGGTCATCATATACCACACAAGGCGGAGGAATGAAGATTGC  
AGCATTACCAACACAGATGAGACCAACTTGCCAGCAGATATTCTAGTTATTTGTCTCT  
CAGGGAACGTTAGCTGACAGGCAGGATGGGTACGTGTCTTCAGAAAGTGGAAGCCACCAC  
CAGTTTGTACATCTTCAGGTGCTGGATTTTTCTTACCACAACATGACAGTAGTGGGACC  
TGCCATATTGACAATAGCAGTGAAGCTGATGTGGAAGCTGCCACAGATCTGTTCTTTGT  
CCGTTTTTGGGATCCACAGGCCCTATGTATTTGAAGGGAAATGTGTATGGCTCAGATCCT  
TTTGAAACATATCATACAGGTTGCAGTCCTGACCAAGAAGACAGTTTAAATGGACCACTAT

**FIG. 25A****SUBSTITUTE SHEET (RULE 26)**

77 / 108

GAGCCCAGTTACATAAAGAAAAAGGAGTGCTACCCATGTTCTCATCCTTCAGAAGAATCC  
TGCGAACGGAGCTTCAGTAATATATCGTGGCCTTCACATGTGAGGAAGCTACTTAACACT  
AGTTACTCTCACAATGAAGGACCTGGAATGAAAAATCTGTGTCTAAACAAGTCCTCTTTA  
GATTTTAGTGCAAATCCAGAGCCAGCGTCGGTTGCCTCGAGTAATTCTTTCATGGGTACC  
TTTGAAAAGCTCTCAGGAGACCTCACCTAGATGCCTATTCAAGCTTTGGACAGCCATCA  
GATTGTCAGCCAAGAGCCTTTTATTTGAAAGCTCATTCTTCCCCAGACTTGGACTCTGGG  
TCAGAGGAAGATGGGAAAGAAAGGACAGATTTTCAGGAAGAAAATCACATTTGTACCTTT  
AAACAGACTTTAGAAAACCTACAGGACTCCAAATTTTCAGTCTTATGACTTGGACACATAG  
ACTGAATGAGACCAAAGGAAAAGCTTAACATACTACCTCAAGTGAACCTTTTATTTAAAAG  
AGAGAGAATCTTATGTTTTTTAAATGGAGTTATGAATTTTAAAAGGATAAAAATGCTTTA  
TTTATACAGATGAACCAAATTACAAAAAGTTATGAAAATTTTATACTGGGAATGATGC  
TCATATAAGAATACCTTTTTTAACTATTTTTTAACTTTGTTTTATGCAAAAAGTATCTT  
ACGTAAATTAATGATATAAATCATGATTATTTTATGTATTTTATAATGCCAGATTTCTT  
TTTATGGAAAATGAGTTACTAAAGCATTTTAAATAATACCTGCCTTGTAACATTTTTTAA  
ATAGAAGTTACTTCATTATATTTTGCACATTATTTAATAAAATGTGTCAATTTGAA

**FIG. 25B**

78 / 108

&gt;&lt;MW: 117438, pI: 5.82, NX(S/T): 12

SEQ ID NO:263

MVDVLLLFSLCLLFHISRPDLSHNRLSFYKASSMSHLQSLREVKLNNNELETIPNLGPVS  
ANITLLSLAGNRIVEILPEHLKEFQSLETLDLSSNNISELQTAFPALQLKYLYLNSNRVT  
SMEPGYFDNLANTLLVLKLNRRNRI SAIPPKMFKLPQLQHLELNRNKIKNVDGLTFQGLGA  
LKSLKMQRNGVTKLMDGAFWGLSNMEILQLDHNNLTEITKGWLYGLLMLQELHLSQNAIN  
RISPDWEFCQKLSELDLTFNHLSRLDDSSFLGLSLLNTLHIGNNRVSYIADCAFRGLSS  
LKTLDLKNNEISWTIEDMNGAFSGLDKLRRLILQGNRIRSITKKAFTGLDALEHLDLSDN  
AIMSLOGNAFSQMKKLQQLHLNTSSLLCDCQLKWLPQWVAENNFQS FVNASCAHPQLKKG  
RSIFAVSPDGFVCDDFPKPQITVQPETQSAIKGSNLSFICSAASSSDSPMTFAWKKDNE  
LHDAEMENYAH LRAQGGEVMEYTTILRLREVEFASEGKYQCVISNHFGSSYSVKAKLTVN  
MLPSFTKTPMDLTIRAGAMARLECAAVGH PAPQIAWQKDGGTDFPAARERRMHVMPEDDV  
FFIVDVKIEDIGVYSCTAQNSAGSISANATLT VLETSPFLRPLLDRTVTKGETAVLQCIA  
GGSPPPKLNWTKDDSPLVVTERHFFAAGNQLLIIVDS DVSDAGKYTCEMSNTLGTERGNV  
RLSVIPTPTCDSPQMTAPSLDDDGWATVGVV IIAVCCVVGTSLVWVVIYHTRRRNEDC  
SITNTDETNPADIPSYLSSQGT LADRQDGYVSSSESGSHHQFVTSSGAGFFLPQHDSSGT  
CHIDNSSEADVEAATDLFLCPFLGSTGPMY LKGNVYGSDPFETYHTGCSPDPRTVLMDHY  
EPSYIKKKECYP CSHPSSESCERSFSNISWPSHVRKLLNTSYSHNEGPGMKNLCLNKSSL  
DFSANPEPASVASSNSFMGTFGKALRRPHL DAYSSFGQPSDCQPRAFYLKAHSSPDLD SG  
SEEDGKERTDFQEEHNICTFKQTL ENYRTPNFQSYDLDT

FIG. 26

79 / 108

SEQ ID NO:265

W22274

1 GGGTCTGTCCATCTTGAGGTATCGTGAACCTTGCCATGTGCAACCTTCGG

W22274

51 GAAATCCCTAACCTCACACCGCTCATAAACTAGATGAGCTGGATCTTTC

W22274

101 TGGGAATCATTATCTGCCATCAGGCCTGGNTCTTCCAGGGTTTGATGC

W22274

151 ACCTTCAAAAACCTGTGGATGATACAGNCCCAGATTCAAGTGATTGANCGG

W22274

SEQ ID NO:266

R55603

SEQ ID NO:264

&lt;DNA36685&gt;

202 ATGCCTTNGACAACCTTCAGTCACTAGTGGAGATCAACCTGGAACACAAT

1 ATGCCTTTGACAACCTTCAGTCACTAGTGGAGATCAACCTGGCACACAAT

1 ATGCCTTTGACAACCTTCAGTCACTAGTGGAGATCAACCTGGCACACAAT

W22274

R55603

&lt;DNA36685&gt;

252 ANTCTAACATTACTGCCTCATGACCTCTTCACTCCCTTGCATCATCTTAG

51 AATCTAACATTACTGCCTCATGACCTCTTCACTCCCTTGCATCATCT-AG

51 AATCTAACATTACTGCCTCATGACCTCTTCACTCCCTTGCATCATCTTAG

W22274

R55603

&lt;DNA36685&gt;

302 AGCGGATACATTTACATCACAACCCTTGGAACCTTGTAACCTGTGACATAC

101 AGCGGATACATTTACATCACAACCCTTGGAACCT-GTAACCT-GTGACATAC

101 AGCGGATACATTTACATCACAACCCTTGGAACCTTGTAACCTGTGACATAC

W22274

R55603

&lt;DNA36685&gt;

352 TTGTGGCTCAAGCTGGTGGATTAAAAGACATGGCCCCCTCGAACACAGGT

151 T-GTGGCTCA-GCTGGTGGAT-AAAAGACATGGCCCCCTCGAACACAGCT

151 TTGTGGCTCAAGCTGGTGGATTAAAAGACATGGCCCCCTCGAACACAGCT

W22274

R55603

&lt;DNA36685&gt;

402 TGTNGTGCCCGGNGTACACTCCTCCCAATCTTAAGGGGGAGGGTCAATG

201 TGTTGTGCCCCGGTGTAACACTCCTCCCAATCTAAAGGGGAGGTACATTGG

201 TGTTGTGCCCCGGTGTAACACTCCTCCCAATCTAAAGGGGAGGTACATTGG

W22274

R55603

&lt;DNA36685&gt;

452 GGGGGCTCGGCCAGATTCCTTGGG

251 AGAGCTCGACCAGAATTACTTCACATGCTATGCTCCGGTGATTGTGGAGC

251 AGAGCTCGACCAGAATTACTTCACATGCTATGCTCCGGTGATTGTGGAGC

FIG. 27A



80 / 108

R55603  
<DNA36685>  
301 CCCCTGCAGACCTCAATGTCACTGAAGGCATGGCAGCTGAGCTGAAATGT  
301 CCCCTGCAGACCTCAATGTCACTGAAGGCATGGCAGCTGAGCTGAAATGT

R55603  
<DNA36685>  
351 TCGGGCCCTCCACATCCCTGACATCTGTATCTTGGGTACTCCAAATGGGA  
351 TCGGGCCCTCCACATCCCTGACATCTGTATCTTGGGTACTCCAAATGGGA

R55603  
<DNA36685>  
401 ACAGTCATGGACACATGGGGGGCGTTACAAAGTTGCGGGTTAGCTGTTGT  
401 ACAGTCATGGACACATGGGGGGCGTTACAAAGTTGCGGGTTAGCTGTTGT

R55603  
<DNA36685>  
451 TCAGTTGATGGTAACGTTTAAATTTTCACAAATGTTAACTGTGGCAAGG  
451 TCAGTTGATGGTAACGTTTAAATTTTCACAAATGTTAACTGTGGCAAGG

FIG. 27B

DNA41388 (3662 bp)									
Scoring parameters: T=12, S=69, S2=36, Matrix: BLOSUM62									
Database: /usr/seqdb/blast/dblast (352,486 entries, 86,705,044 aa)									
Sequences producing High-scoring Segment Pairs:									
		glial cell membrane glycoprotein LIG-1 pr...	Frame	Score	Match	Pct			
1	A58532		+2	2517	482	60			
2	CEL21D12_3	T21D12.9a - Caenorhabditis elegans	+2	864	222	33			
3	CEL21D12_1	T21D12.9b - Caenorhabditis elegans	+2	864	222	33			
4	JC6128	insulin-like growth factor binding comple...	+2	363	121	30			
5	ALS_MOUSE	Insulin-like growth factor binding proteol...	+2	363	121	30			
6	GEN11209	18 wheeler - Drosophila melanogaster	+2	350	118	32			
7	DROWHEELER_118w	Drosophila melanogaster	+2	350	118	32			
8	JC5239	insulin-like growth factor acid-labile ch...	+2	348	98	33			
9	S83462_1	ALS - Papio	+2	348	98	33			
10	CELC56E6_6	C56E6.6 - Caenorhabditis elegans	+2	348	108	29			
>1 A58532 glial cell membrane glycoprotein LIG-1 precursor - mouse (1091 aa)									
Score = 2517 (886.0 bits), Expect = 1.6e-263, Sum P(2) = 1.6e-263									
Identities = 482/802 (60%), Positives = 619/802 (77%), at 122,76, Frame = +2									
DNA41388	122	DLSHNRISFIKASSMSHLQSLREVKLNNELETIPNLGPVSANITLLSLAGNRIVEILPE							
		..***** * ... * ..** **..*** **..** * ..** * ..**							
A58532	76	NLSYNRLSEIDSAAFEDLTNLQEVYVLSNELTAIPSLGTASIGVVSFLQHNKILSVDSGS							
DNA41388	302	HLKEFQSLETLDLSSNNISELQTA-FP-ALQLKYLVLNSNRVTSMEPGYFDNLANTLLVL							
		** . *** *****..*.... ** *... * * ***..* * * * . ** *							
A58532	136	QLKSYLSLEVLDLSSNNITEIRSSCFPNGLRIRLNLASNRISILESGAFDGLSRSLTL							
DNA41388	476	KLNRNRISAIPPKMFKLPLQLHLELNRNKIKNVVDGLTFQGLGALKSLKMQRNGVTKLMDG							
		..****. . * * ****. * ..****. * ..***** * ..**** * ..** *							
A58532	196	RLSKNRITQLPVKAFKLPRLTQLDLNRNRIRLIEGLTFQGLDSLVLRLQRNNISRLTDG							
DNA41388	656	AFWGLSNMEILQLDHNLTETITKGWLYGLLMQLQLHLSQNAINRISPDWAFECQKLSELD							
		***** * . * *...* * . * **** * ..**** * ..**** * ..**** *							
A58532	256	AFWGLSKMHVLHLEYNLSLVENSGSLYGLTALHQLHLSNNISIRIQRDGWSFCQKLHELI							

FIG. 28A

**SUBSTITUTE SHEET (RULE 26)**



84 / 108

SEQ ID NO:270

MARPGPGVLGAPRLAPRLLLWLLLLLLQWPESAGAQAARPRAPCAAACCTCAGNSLDCSGRG  
LATLPRDLPSWTRSLNLSYNRLSEIDSAAFEDLTNLQEVYLNSELTAIPSLGTASIGVV  
SLFLQHNKILSVDGSQLKSYLSLEVLDLSSNNITEIRSSCFPNGLRIRELNLASNRISIL  
ESGAFDGLSRSLTLRLSKNRITQLPVKAFKLPRLTQLDLNRNRIRLIEGLTFQGLDSLE  
VLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVNSGSLYGLTALHQLHLSNNSISRI  
QRDGWSFCQKLHELILSFNNLTRLDEESLAELSSLSILRLSHNAISHIAEGAFKGLKSLR  
VLDLDHNEISGTIEDTSGAFTGLDNLSKLTFLGNKIKSVAKRAFSGLESLEHLNLGENAI  
RSVQFDAFAKMKNLKELYISSESFCLDCQLKWLPWPWLMGRMLQAFVTATCAHPESLKGQS  
IFSVL PDSFVCDDFPKPQIIITQPETTMVVGKDIRFTCSAASSSSSPMTFAWKKDNEVLA  
NADMENFAHVRAQDGEVMEYTTILHLRHVTFGHEGRYQCIITNHFGSTYSHKARLTVNVL  
PSFTKIPHDIAIRTGTTARLECAATGHPNPQIAWQKDGGTDFPAARERRMHVMPDDDDVFF  
ITDVKIDDMGVYSCTAQNSAGSVSANATLTVLETPSLAVPLEDRVVTVGETVAFQCKATG  
SPTPRITWLKGGRP LSLTERHHFTPGNQLLVVQNV MIDDAGRYTCEMSNPLGTERAHSQ L  
SILPTPGCRKDGTTVGIFTIAVVC SIVLTSLVWVCIIYQTRKKSEEYSVTNTDETIVPPD  
VPSYLSSQGTLSDRQETVVRTEGGHQANGHIESNGVCLRDPSLFPEVDIHSTTCRQPKLC  
VGYTREPWKVTEKADRTAAPHTTAHSGSAVCSDCSTD TAYHPQPVPRDSGQPGTASSQEL  
RQHDREYSPHPYSGTADGSHTLSGGSLYPSNHDRILPSLKNKAASADGNGDSSWTLAKL  
HEADCIDLKPSPTLASGSPELMEDAISTEAQHLLVSNHGLPKACDSSPESVPLKGQITGK  
RRG PLLAPRS

FIG. 29A

85 / 108

<71-94/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1\}  
 <95-117/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2\}  
 <118-141/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3\}  
 <142-165/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4\}  
 <166-189/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5\}  
 <191-213/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6\}  
 <214-237/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7\}  
 <238-261/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8\}  
 <262-285/domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9\}  
 <286-309/domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10\}  
 <310-333/domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11\}  
 <334-357/domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12\}  
 <358-381/domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13\}  
 <385-408/domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14\}  
 <409-432/domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15\}

FIG. 29B

86 / 108

inseq working...  
<2228990 PROSNOT16 gl545806 House Mouse; musculus domesticus  
mRNA fo gb101rod 24 -1>  
TTGGGGTATACAGCTGCACAGCTCAGAACAGTGCAGGAAGTATTTTCAGCAAATGCAACTC  
TGACTGTCCTAGAAACACCATCATTTTTTGC GGCCACTGTTGGACCGAACTGTAACCAAGG  
GAGAAACAGCCGTCCTACAGTGCATTGCTGGAGGAAGCCCTCCCCCTAAACTGAACTGGA  
CCAAAGATGATAGCCCATTTGGTGGTAACCGAGAGGCACTTTTTTGCAGCAGGC

SEQ ID NO: 76

**FIG. 30A**

/usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA36749 (233 bp)  
 Scoring parameters: T=12, S=58, S2=31, Matrix: BLOSUM62  
 Database: /usr/seqdb/blast/dbblast (321,232 entries, 78,212,008 aa)

Sequences producing High-scoring Segment Pairs:		Frame		Score	Match	Pct
1	PTPF_HUMAN Lar protein precursor - homo sapiens	+3		135	31	43
2	S46216 leukocyte antigen-related protein precursor -	+3		135	31	43
3	PN0568 connectin 3B - chicken (fragment)	+3		134	28	37
4	GEN13581 Muscle-specific kinase (MuSK) - human	+3		132	26	39
5	HSU48959_1 myosin light chain kinase - Homo sapiens	+3		130	29	40
6	HSTITINN2_1 elastic titin - Homo sapiens	+		130	24	33
7	CAU55211_1 L1-like cell adhesion molecule antigen E5...	+3		129	26	37
8	S46224 peroxidasin - fruit fly (Drosophila sp.)	+3		129	25	38
9	DMU11052_1 peroxidasin precursor - Drosophila melano...	+3		129	25	38
10	B48758 protein-tyrosine-phosphatase (EC 3.1.3.48), r	+3		125	30	42

FIG. 30B

```
><DNA36749: 2228990 PROSNOT16 g1545806 House Mouse; musculus domesticus mRNA fo gb98rod 24 -1>
TT
><ORF {trans=1-s, dir=f, res=1}>
GGGGTAT
><36749.f1 {underline=1-24, dir=f}>
ACAGCTGCACAGCTCAGAACAGTGCAGGAAGTATTTTCAGCAATGCAACTCTGACTGTCC
TAGAAACACCATCATTTT SEQ ID NO:277
><36749.p1 {underline=1-50, dir=f}>
GCGGCCACTGTTGGACCGAACTGTAACCAAGGAGAAACAGCCGTCCTACAGTGCAATTGC
TGGAGGAAGCCCTCCCCCTAAACTGAACTGGACCAAGATGATAGCCCCATTGGTGGTAAC
CGAGAGGCA SEQ ID NO:278
><36749.r1 {underline=1-24, dir=b}>
CTTTTTTGCAGCAGGC SEQ ID NO:279
```

FIG. 30C



88 / 108

GGAACCGAATCTCAGCTA SEQ ID NO:271

CATTCCTCAGTATAAAAATTTTC SEQ ID NO:275

GGGTCTTGGTGAATGAGG SEQ ID NO:276

CCTAAACTGAACTGGACCA SEQ ID NO:272

GGCTGGAGACACTGAACCT SEQ ID NO:273

ACAGCTGCACAGCTCAGAACAGTG SEQ ID NO:274

GCGGCCACTGTTGGACCGAACTGTAACCAAGGGAGAAACAGCCGTCCTAC SEQ ID NO:278

GTGCCTCTCGGTTACCACCAATGG SEQ ID NO:277

**FIG. 31**

89 / 108

SEQ ID NO:279

GGGGAGAGGAATTGACCATGTAAAAGGAGACTTTTTTTTTTGGTGGTGGTGGCTGTTGGG  
TGCCTTGCAAAAATGAAGGATGCAGGACGCAGCTTTCTCCTGGAACCGAACGCAATGGAT  
AAACTGATTGTGCAAGAGAGAAGGAAGAACGAAGCTTTTTCTTGTGAGCCCTGGATCTTA  
ACACAAATGTGTATATGTGCACACAGGGAGCATTCAAGAATGAAATAAACCAGAGTTAGA  
CCCGCGGGGGTTGGTGTGTTCTGACATAAATAAATAATCTTAAAGCAGCTGTTCCCTCC  
CCACCCCCAAAAAAGGATGATTGGAAATGAAGAACCGAGGATTCACAAAGAAAAAAGT  
ATGTTTCATTTTTCTCTATAAAGGAGAAAGTGAGCCAAGGAGATATTTTTGGAATGAAAAG  
TTTGGGGCTTTTTTAGTAAAGTAAAGAACTGGTGTGGTGGTGTTCCTTTCTTTTTGAA  
TTTCCCAAGAGGAGAGGAAATTAATAATACATCTGCAAAGAAATTCAGAGAAGAAAA  
GTTGACCGCGGCAGATTGAGGCATTGATTGGGGGAGAGAAACCAGCAGAGCACAGTTGGA  
TTTGTGCCTATGTTGACTAAAATTGACGGATAATTGCAGTTGGATTTTTCTTCATCAACC  
TCCTTTTTTTTTAAATTTTTATTCCTTTGGTATCAAGATCATGCGTTTTCTCTGTTCTT  
AACCACCTGGATTTCCATCTGGATGTTGCTGTGATCAGTCTGAAATACAACCTGTTTGAAT  
TCCAGAAGGACCAACACCAGATAAATTATGA

SEQ ID NO:280

ATGTTGAACAAGATGACCTTACATCCACAGCAGATAATGATAGGTCCTAGGTTTAACAGG  
GCCCTATTTGACCCCCTGCTTGTGGTGCTGCTGGCTCTTCAACTTCTTGTGGTGGCTGGT  
CTGGTGCGGGCTCAGACCTGCCCTTCTGTGTGCTCCTGCAGCAACCAGTTCAGCAAGGTG  
ATTTGTGTTTCGAAAAACCTGCGTGAGGTTCCGGATGGCATCTCCACCAACACACGGCTG  
CTGAACCTCCATGAGAACCAAATCCAGATCATCAAAGTGAACAGCTTCAAGCACTTGAGG  
CACTTGGAATCCTACAGTTGAGTAGGAACCATATCAGAACCATTGAAATTGGGGCTTTC  
AATGGTCTGGCGAACCTCAACACTCTGGAACCTTTTGACAATCGTCTTACTACCATCCCG  
AATGGAGCTTTTGTATACTTGTCTAACTGAAGGAGCTCTGGTTGCGAAACAACCCCAT  
GAAAGCATCCCTTCTTATGCTTTTAACAGAATTCCTTCTTTGCGCCGACTAGACTTAGGG  
GAATTGAAAAGACTTTCATACATCTCAGAAGGTGCCTTTGAAGGTCTGTCCAACCTGAGG  
TATTTGAACCTTGCCATGTGCAACCTTCGGGAAATCCCTAACCTCACACCGCTCATAAAA  
CTAGATGAGCTGGATCTTTCTGGGAATCATTTATCTGCCATCAGGCCTGGCTCTTCCAG  
GGTTTGATGCACCTTCAAAAACCTGTGGATGATACAGTCCCAGATTCAAGTGATTGAACGG  
AATGCCTTTGACAACCTTCAGTCACTAGTGGAGATCAACCTGGCACACAATAATCTAACA  
TTACTGCCTCATGACCTCTTCACTCCCTTGATCATCTAGAGCGGATACATTTACATCAC  
AACCCTTGGAACCTGTAACCTGTGACATACTGTGGCTCAGCTGGTGGATAAAAGACATGGCC  
CCCTCGAACACAGCTTGTGTGCGCCGGTGTAACTCCTCCCAATCTAAAGGGGAGGTAC  
ATTGGAGAGCTCGACCAGAATTACTTCACATGCTATGCTCCGGTGATTGTGGAGCCCCCT  
GCAGACCTCAATGTCACTGAAGGCATGGCAGCTGAGCTGAAATGTGCGGCCTCCACATCC  
CTGACATCTGTATCTTGGATTACTCCAAATGGAACAGTCATGACACATGGGGCGTACAAA  
GTGCGGATAGCTGTGCTCAGTGATGGTACGTTAAATTTACAAATGTAACGTGTGCAAGAT  
ACAGGCATGTACACATGTATGGTGAGTAATTCGTTGGGAATACTACTGCTTCAGCCACC  
CTGAATGTTACTGCAGCAACCACTACTCCTTTCTTACTTTTCAACCGTCACAGTAGAG  
ACTATGGAACCGTCTCAGGATGAGGCACGGACCACAGATAACAATGTGGGTCCCCTCCA  
GTGGTCGACTGGGAGACCACCAATGTGACCACCTCTCTCACACCACAGAGCACAAAGTGC  
ACAGAGAAAACCTTCACCATCCCAGTGACTGATATAAACAGTGGGATCCCAGGAATTGAT  
GAGGTCATGAAGACTACCAAAATCATCATTGGGTGTTTTGTGGCCATCACACTCATGGCT  
GCAGTGATGCTGGTCATTTTCTACAAGATGAGGAAGCAGCACCATCGGCAAAACCATCAC  
GCCCCAACAAAGGACTGTTGAAATTATTAATGTGGATGATGAGATTACGGGAGACACACCC  
ATGGAAAGCCACCTGCCCATGCCTGCTATCGAGCATGAGCACCTAAATCACTATAACTCA  
TACAAATCTCCCTTCAACCACACAACAACAGTTAACACAATAAATTCAATACACAGTTCA  
GTGCATGAACCGTTATTGATCCGAATGAACCTCTAAAGACAATGTACAAGAGACTCAAATC  
TAAAACATTTACAGAGTTACAAAAACAAACAATCAAAAAAAGACAGTTTATTAAAAA  
TGACACAAATGACTGGGCTAAATCTACTGTTTCAAAAAAGTGTCTTTACAAAAAACAAA  
AAAGAAAAGAAATTTATTTATTAATAATTCTATTGTGATCTAAAGCAGACAAAA

FIG. 32

SUBSTITUTE SHEET (RULE 26)

90 / 108

><MW: 71950, pI: 7.12, NX(S/T): 10  
(MLNKMTLHPQQIMIGPRFNRAFDPLLVLALLQLLVVAGLVRA)QTCPSVCSCSNQFSKV  
ICVRKQLREVDPDGISTNTRLNLHENQIQI<sup>X</sup>IKVNSFKHLRHLEILQLSRNHIRTIEIGAF  
NGLANLNTLELFDNRLTTIP<sup>X</sup>NGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRDLG  
ELKRLSYISEGAFEGLSNLRYNLAMCNLREIPNLTPLIKLDLDELDSGNHLSAIRPGSFQ  
GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH  
NPWNCNC<sup>X</sup>DILWLSWWIKD<sup>X</sup>MAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP  
ADLNVTEGMAAELKCRAS<sup>X</sup>TSLSVSWITPNGTVMTHGAYKVRIAVLS<sup>X</sup>DGTLNFTNVTVQD  
TGM<sup>X</sup>YTCMVSN<sup>X</sup>SVGNTTASATLNVTAATTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP  
VVDWETTNVTTS<sup>X</sup>SLTPQSTRSTEKFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA  
AVML<sup>X</sup>VIFYKMRKQHHRQNH<sup>X</sup>HAPTRIVEIINVDDEITGDTPMESHLPMPAIEHEHLNH<sup>X</sup>YNS  
YKSPFNHTTTVNTINSIHSSVHEPLLRMNSKDNVQETQI

SEQ ID NO:281

FIG. 33

Sequences producing High-scoring Segment Pairs:	glial cell membrane glycoprotein	LIG-1	prec..	+2	353	119	28
1 A58532							
2 D86983_1	KIAA0230 - Homo sapiens			+2	337	97	30
3 JC5239	insulin-like growth factor acid-labile chain			+2	325	95	36
4 ALS_PAPPA	Insulin-like growth factor binding protein			+2	325	95	36
5 ALS_HUMAN	Insulin-like growth factor binding protein			+2	312	92	34
6 P_R85888	WD-40 domain-contg. insulin-like growth factor			+2	312	92	34
7 PGS2_HUMAN	Bone proteoglycan 11 precursor - homo sapiens			+2	305	85	33
8 P_R89439	Human recombinant decorin - Homo sapiens.			+2	305	85	33
9 P_R42260	Mature decorin PT-65 - unknown			+2	305	85	33
10 P_R42267	Decorin sequence PT-78 (N-terminal to half			+2	305	85	33

>1 A58532 glial cell membrane glycoprotein LIG-1 precursor - mouse (1091 aa)

Score = 353 (124.3 bits), Expect = 1.5e-27, P = 1.5e-27

Identities = 119/418 (28%), Positives = 200/418 (47%), at 1052, 218, Frame = +2

DNA40981 1052 LNLHENQIQI I K V N S F K H L R H L E I L Q L S R N H R T I E I G A F N G L A N L N T L E U F D N R L T T I P  
 \* \* \* \* \* . \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \*

SEQ ID NO:282

210 LDLNRRIRLIEGLTFOGLDSLEVLRLORNNISRLTDGAFWGLSKMHVLHLEYNSLVEVN

DNA40981 1232 NGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRRLDLGELKRLSYISEGAFEGLSNLR

A50532 278 SGLYGLTALHOLHLSNNSISRIORDGWSFCOKLHELILS-FNNLTRLDEESLAEELSSLS

DNA40981 1412 YLNAMCNLREIPN--LTPLIKLDELDSGNHLSAI---RPGSFOGLMHLOKLMWIOSOI

A58532 337 ILRLSHNAISHIAEGAFKGLKSLRVLDLDPHNEISGTIEDTSGAFTGLDNL SKLTLFGNKI

DNA40981 1577 OVIERNAFDNL0SLVEINLAHNNLTLLPHDLEFTPPLHHLERITHLHHNPWNCNCDIJWLSWW

**FIG. 34A**

DNA40981 1757 I--KDMAPNTACCARCNTPPNLKGRYIGELDQNYFTCY--APVIVEPPADLNVTGMA  
 . . . \* \* \* \* \* . . . \* \* \* \* \* . . . \* \* \* \* \*  
 A58532 457 LMGRMLQAFVTATCAH--PESLKGQSIFSVPDSFVCDPFPKQIITQPETMAVVGKD  
 DNA40981 1922 AELKCRASTSLTS---VSWITPNGTV---MTHGAYKVRIA--VLSDGT-LNFTNVTVQ  
 \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \*  
 A58532 514 IRFTCSAASSSSPMTFAWKDNEVLANADMENFAH-VRAQDGEVMEYTTILHLRHVTFG  
 DNA40981 2069 DTGMYTCMVNSVGNNTASATLNVTAATTPFSYFS-TVTVETMEPSQDEARTDNNVGP  
 \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \*  
 A58532 573 HEGRYQCIITNHFSTY-SHKARLTNVNLPSTFKIPHDIAIRGTGTARLECAATGH---P  
 DNA40981 2246 TPVVDWE  
 \* \* \* \* \*  
 A58532 629 NPQIAWQ

Score = 251 (88.4 bits), Expect = 4.0e-25, Sum P(2) = 4.0e-25  
 Identities = 77/222 (34%), Positives = 110/222 (49%), at 1052,122, Frame = +2

SEQ ID NO:283  
 DNA40981 1052 LNLHENQIQIIVKNSFKHLRHLLEILQLSRNHRTIEIGAP-NGLANLNTLELFDNRLTTI  
 \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \*  
 A58532 122 LFLQHNKILSVDSGSQLKSYSLEVLDLSSNNITEIRSSCFPNGL-RIRELNLSNRISIL  
 DNA40981 1229 PNGAFVYLSK-LKELWLRNPNPIESIPSYAFNRIPSLRRDLGELKRLSYISEGAFEGLSN  
 . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \*  
 A58532 181 ESGAFDGLSRSLTLRLSKNRITQLPVKAF-KLPRLTQLDLNR-NRIRLIEGLTFQGLDS  
 DNA40981 1406 LRYLNLMCNLREIPN--LTPLIKLDELDSGNHLSAIRPGSFQGLMHQLKLMWMIQSQIQ  
 \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \*  
 A58532 239 LEVLRQRNNISRLTDGAFWGLSKMNVHLLEYNSLVEVNSGSLYGLTALHQLHLSNNSIS  
 DNA40981 1580 VIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLLHN  
 \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \* . \* \* \* \* \*  
 A58532 299 RIQRDGSWFCQKLHELILSFNNLRLDEESLAELSSLSILRLSHN

FIG. 34B

93 / 108

Score = 237 (83.4 bits), Expect = 1.3e-23, Sum P(2) = 1.3e-23  
 Identities = 86/290 (29%), Positives = 147/290 (50%), at 845,8, Frame = +2

DNA40981 845 IMIGPRENRAFDPLLVLLALQLLVAGL-VRAQT-CPSVCSCSNQFSKVICVRKNLRE  
 .. \*\* \* ..\*\* \*\* \*\* \* ..\*\* \* ..\*\* \* ..\*\* \* ..\*\* \*  
 SEQ ID NO:284 A58532 8 VLGAAPRLAPRL--LWLLLLLLQWPESAGAQAPRACAACTCAG--NSLDCSGRGLAT

DNA40981 1019 VPDGISTNTRLNLHENQIQIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTL  
 . \* .. \*\* \*\*\* \*\* \* ..\*\* \* ..\*\* \* ..\*\* \* ..\*\* \* ..\*\* \*  
 A58532 64 LPRDLPSWTRSLNLSYNRLSEIDSAAFEDLTNLQEVYLSNELTAIPS---LGTASIGVV

DNA40981 1199 ELF--DNRLTTIPNGAFV-YLSKCLKELWLRNPNPIESIPSYAFNRIPSLRRDLGELKRLS  
 \*\* \* .. \*\* \*\*\* \*\* \* ..\*\* \* ..\*\* \* ..\*\* \* ..\*\* \* ..\*\* \*  
 A58532 121 SLFLQHNKILSVDSQLKSYLS-LEVLDLSSNNITEIRSSCFPNGLRIRELNLAS-NRIS

DNA40981 1370 YISEGAFEGLS-NLRYNLAMCNLREIP-NLTPLIKDELDSLGNHLSAIRPGSFQGLMH  
 . \*\*\*.\*\*\* \* \* \* \* . . . . \* \* \* \* \* \* \* \* \* \* \* \* \*  
 A58532 179 ILESGAFDGLSRSLTLRLSKNRITQLPVKAFKLPRLTQLDLNRNRIRLIEGLTFQGLDS

DNA40981 1544 LQKLWMIQSIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLLHN  
 \* \* \* .. \*  
 A58532 239 LEVLRQRNNISRLTDGAFWGLSKMHVHLHLEYNLSVEVNSGSLYGLTALHQLHLSNN

Score = 194 (68.3 bits), Expect = 5.5e-19, Sum P(2) = 5.5e-19  
 Identities = 57/200 (28%), Positives = 102/200 (51%), at 1049,241, Frame = +2

DNA40981 1049 LLNLHENQIQIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTI  
 . \* \* \* \* . . . \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*  
 SEQ ID NO:285 A58532 241 VLRLQRNNISRLTDGAFWGLSKMHVHLHLEYNLSVEVNSGSLYGLTALHQLHLSNNSISRI  
 DNA40981 1229 PNGAFVYLSKCLKELWLRNPNPIESIPSYAFNRIPSLRRDLGELKRLSYISEGAFEGLSNL  
 . . . \*\* \*\* \* \* \* \* . . . . \* \* \* \* \* \* \* \* \* \* \* \* \*  
 A58532 301 QRDGWSFCQKLHELILSFNNLRLDEESLAEELSSLSILRLSH-NAISHIAEGAFKGLKSL

FIG. 34C

DNA40981 1409 RYLNLCNLR-EIPN-...-LTPLIKDELDSLGNHLSAIRPGSFQGLMHLQKLWMIQSQ  
 \* \* \* \* \*  
 A58532 360 RVLDLHNEISGTIEDTSGAFTGLDNLKLTFLGNKIKSVAKRAFGSLESLEHNLNGENA  
 \* \* \* \* \*  
 DNA40981 1574 IQVIERNAFDNLQSLVEINLA  
 \* \* \* \* \*  
 A58532 420 IRSVQFDAFAKMKNLKELYIS  
 \* \* \* \* \*

Score = 163 (57.4 bits), Expect = 1.1e-15, Sum P(2) = 1.1e-15  
 Identities = 55/152 (36%), Positives = 78/152 (51%), at 1025,303, Frame = +2

DNA40981 1025 DGISTNTRL--LNLHENQIQIHKVNSFKHLRHLLEILQLSRNHRTIEIGAFNGLANLNTL  
 \* \* \* \* \*  
 A58532 303 DGWFCQKLHELILSFNNLTRLDEESLAEELSSILRLSHNAISHIAEGAFKGLKSLRVL  
 \* \* \* \* \*  
 DNA40981 1199 ELFDNRLT-TIPN--GAFVYLSKLKELWLRNPNIESIPSYAFNRIPSLRRLDLGELKRRLS  
 \* \* \* \* \*  
 A58532 363 DLDHNEISGTIEDTSGAFTGLDNLKLTFLGNKIKSVAKRAFGSLESLEHNLNGE-NAIR  
 \* \* \* \* \*  
 DNA40981 1370 YISEGAFEGLSNLR--YLN---LAMCNLREIP  
 \* \* \* \* \*  
 A58532 422 SVQFDAFAKMKNLKELYISSEFLCDCQKWL  
 \* \* \* \* \*

Score = 135 (47.5 bits), Expect = 4.0e-25, Sum P(2) = 4.0e-25  
 Identities = 44/141 (31%), Positives = 67/141 (47%), at 1871,601, Frame = +2

DNA40981 1871 PVIVEPPADLNVTEGMAAELKCRASLT-SVSWITPNGTVMTHGAYKVRIAVLSDGTLN  
 \* \* \* \* \*  
 A58532 601 PSFTKIPHDIAIRGTGTARLECAATGHPNPQIAWQKGGTDFP-AARERRMHVMPDDDDVF  
 \* \* \* \* \*  
 DNA40981 2048 F-TNVTVDGTGMYTCMVNSVGNNTTASATLNV--TAATTTFFSYFSTVTETMEPSQDEA  
 \* \* \* \* \*  
 A58532 660 FITDVKIDDMGVYCTAQNAGSVSANATLTVLETPSLAVPLED-RVVTVG--ETVAFQC  
 \* \* \* \* \*

FIG. 34D





96 / 108

/usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA36685 (499 bp)  
 Scoring parameters: T=12, S=61, S2=34, Matrix: BLOSUM62  
 Database: /usr/seqdb/blast/dblast (318,238 entries, 77,505,313 aa)

Sequences producing High-scoring Segment Pairs:	Frame	Score	Match	Pct
1 S46224 peroxidase - fruit fly (Drosophila sp.)	+3	128	35	28
2 DMU11052_1 peroxidase precursor - Drosophila melano...	+3	128	35	28
3 SLIT_DROME Slit protein precursor - drosophila melan...	+3	115	31	34
4 P_R25079 Drosophila SLIT protein involved in axon pa...	+3	109	29	32
5 GPV_HUMAN Platelet glycoprotein v precursor - homo s...	+3	88	21	35
6 P_R71294 Human glycoprotein V - Homo sapiens.	+3	88	21	35
7 HSU59632_2 platelet glycoprotein Ib beta chain - Hom...	+3	77	31	36
8 GPBB_HUMAN Platelet glycoprotein Ib beta chain precu...	+3	77	31	36
9 DROSGS4C1_1 Sgs4 - Drosophila melanogaster	+ +3	58	13	41
10 DROSGS4H1_1 Sgs4 - Drosophila melanogaster	+ +3	58	13	41

FIG. 35A

```

>1 S46224 peroxidase - fruit fly (Drosophila sp.) (1535 aa)
Score = 128 (45.1 bits), Expect = 4.2e-06, P = 4.2e-06
Identities = 35/126 (27%), Positives = 54/126 (42%), at 6,144, Frame = +3

DNA36685      6 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLRADTFTSQPLEL-VTCDILVAQAGGLKDM
      **** * . . . * * * * * * * * * * * * . . . * * *
SEQ ID NO:293  S46224    144 FDNLPRLNRLIMYNNKLTQLPVDGFNRLNNLKRLLDGNDAIDIDCNCGVYSLWRRWHLDV
      * * * * * * * * * * * * * * * * * * * * * * * *
DNA36685    183 APSNTACCARCNTPPNLKGRYIGELDNQYFTCYAPVIVEPPADLNVTEGMAAELKC--SG
      * * * * * * * * * * * * * * * * * * * * * * * *
S46224      204 QRQLVSISLTCAAPQMLQNQGFSLSGEHHFKCAKQPFLVAPQDAQVAAGEQVELSCEVTG
      * * * * * * * * * * * * * * * * * * * * * * * *

DNA36685    357 LHIPDI
      ** * *
S46224      264 LHRPQI

>2 DMU11052_1 peroxidase precursor - Drosophila melanogaster (1535 aa)
Score = 128 (45.1 bits), Expect = 4.2e-06, P = 4.2e-06
Identities = 35/126 (27%), Positives = 54/126 (42%), at 6,144, Frame = +3

DNA36685      6 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLRADTFTSQPLEL-VTCDILVAQAGGLKDM
      **** * . . . * * * * * * * * * * * * . . . * * *
SEQ ID NO:294  DMU11052_1  144 FDNLPRLNRLIMYNNKLTQLPVDGFNRLNNLKRLLDGNDAIDIDCNCGVYSLWRRWHLDV
      * * * * * * * * * * * * * * * * * * * * * * * *
DNA36685    183 APSNTACCARCNTPPNLKGRYIGELDNQYFTCYAPVIVEPPADLNVTEGMAAELKC--SG
      * * * * * * * * * * * * * * * * * * * * * * * *
DMU11052_1  204 QRQLVSISLTCAAPQMLQNQGFSLSGEHHFKCAKQPFLVAPQDAQVAAGEQVELSCEVTG
      * * * * * * * * * * * * * * * * * * * * * * * *

DNA36685    357 LHIPDI
      ** * *
DMU11052_1  264 LHRPQI

```

98 / 108

GCCTTTGACAACCTTCAGTCACTAGTGG SEQ ID NO:295

TACTGCCTCATGACCTCTTCACTCCCTTGCATCATCTTAGAGCGG SEQ ID NO:297

CCCCATGTGTCCATGACTGTTCCC SEQ ID NO:296

**FIG. 36**

99 / 108

SEQ ID NO:298

AGCCGACGCTGCTCAAGCTGCAACTCTGTTGCAGTTGGCAGTTCTTTTCGGTTTCCCTCC  
TGCTGTTTGGGGGCATGAAAGGGCTTCGCCGCCGGGAGTAAAAGAAGGAATTGACCGGGC  
AGCGCGAGGGAGGAGCGCGCACGCGACCGCGAGGGCGGGCGTGCACCCTCGGCTGGAAGT  
TTGTGCCGGGCCCCGAGCGCGCGCCGGCTGGGAGCTTCGGGTAGAGACCTAGGCCGCTGG  
ACCGCG

SEQ ID NO:299

ATGAGCGCGCCGAGCCTCCGTGCGCGCGCCGCGGGGTGGGGCTGCTGCTGTGCGCGGTG  
CTGGGGCGCGCTGGCCGGTCCGACAGCGGCGGTGCGGGGAACTCGGGCAGCCCTCTGGG  
GTAGCCGCCGAGCGCCCATGCCCACTACCTGCCGCTGCCTCGGGGACCTGCTGGACTGC  
AGTCGTAAGCGGCTAGCGCGTCTTCCCGAGCCACTCCCGTCCTGGGTCGCTCGGCTGGAC  
TTAAGTCACAACAGATTATCTTTCATCAAGGCAAGTTCCATGAGCCACCTTCAAAGCCTT  
CGAGAAGTGAACTGAACAACAATGAATTGGAGACCATTCCAAATCTGGGACCAGTCTCG  
GCAAATATTACACTTCTCTCCTTGGCTGGAAACAGGATTGTTGAAATACTCCCTGAACAT  
CTGAAAGAGTTTCAGTCCCTTGAACTTTGGACCTTAGCAGCAACAATATTTTCAGAGCTC  
CAAATGCAATTTCCAGCCCTACAGCTCAAATATCTGTATCTCAACAGCAACCGAGTCACA  
TCAATGGAACCTGGGTATTTTGACAATTTGGCCAACACACTCCTTGTGTTAAAGCTGAAC  
AGGAACCGAATCTCAGCTATCCCACCCAAGATGTTTAAACTGCCCCAACTGCAACATCTC  
GAATTGAACCGAAACAAGATTAAAAATGTAGATGGACTGACATTCCAAGGCCTTGGTGCT  
CTGAAGTCTCTGAAAATGCAAAGAAATGGAGTAACGAAACTTATGGATGGAGCTTTTTGG  
GGGCTGAGCAACATGGAAATTTTGACAGCTGGACCATAACAACCTAACAGAGATTACCAAA  
GGCTGGCTTTACGGCTTGCTGATGCTGCAGGAACTTCATCTCAGCCAAAATGCCATCAAC  
AGGATCAGCCCTGATGCCTGGGAGTTCTGCCAGAAGCTCAGTGAGCTGGACCTAACTTTC  
AATCACTTATCAAGGTTAGATGATTCAAGCTTCCTTGGCCTAAGCTTACTAAATACACTG  
CACATTGGGAACAACAGAGTCAGCTACATTGCTGATTGTGCCTTCCGGGGGCTTTCCAGT  
TTAAAGACTTTGGATCTGAAGAACAATGAAATTTCTGGACTATTGAAGACATGAATGGT  
GCTTCTCTGGGCTTGACAACTGAGGCGACTGATACTCCAAGGAAATCGGATCCGTTCT

FIG. 37A

100 / 108

ATTACTAAAAAGCCTTCACTGGTTTGGATGCATTGGAGCATCTAGACCTGAGTGACAAC  
GCAATCATGTCTTTACAAGGCAATGCATTTTACAAATGAAGAACTGCAACAATTGCAT  
TTAAATACATCAAGCCTTTTGTGCGATTGCCAGCTAAAATGGCTCCCACAGTGGGTGGCG  
GAAAACAACCTTTCAGAGCTTTGTAAATGCCAGTTGTGCCCATCCTCAGCTGCTAAAAGGA  
AGAAGCATTTTTGCTGTTAGCCCAGATGGCTTTGTGTGTGATGATTTTCCCAAACCCAG  
ATCACGGTTCAGCCAGAAACACAGTCGGCAATAAAAGGTTCCAATTTGAGTTTCATCTGC  
TCAGCTGCCAGCAGCAGTGATTCCCCAATGACTTTTGCTTGGAAAAAGACAATGAACTA  
CTGCATGATGCTGAAATGGAAAATTATGCACACCTCCGGGCCCAAGGTGGCGAGGTGATG  
GAGTATACCACCATCCTTCGGCTGCGCGAGGTGGAATTTGCCAGTGAGGGGAAATATCAG  
TGTGTCATCTCCAATCACTTTGGTTCATCCTACTCTGTCAAAGCCAAGCTTACAGTAAAT  
ATGCTTCCCTCATTACCAAGACCCCCATGGATCTCACCATCCGAGCTGGGGCCATGGCA  
CGCTTGGAGTGTGCTGCTGTGGGGCACCCAGCCCCCAGATAGCCTGGCAGAAGGATGGG  
GGCACAGACTTCCCAGCTGCACGGGAGAGACGCATGCATGTGATGCCCCGAGGATGACGTG  
TTCTTTATCGTGGATGTGAAGATAGAGGACATTGGGGTATACAGCTGCACAGCTCAGAAC  
AGTGCAGGAAGTATTTTTCAGCAAATGCAACTCTGACTGTCCTAGAAACACCATCATTTTTG  
CGGCCACTGTTGGACCGAACTGTAACCAAGGGAGAAACAGCCGTCCTACAGTGCATTGCT  
GGAGGAAGCCCTCCCCCTAAACTGAACTGGACCAAAGATGATAGCCCATTTGGTGGTAACC  
GAGAGGCACTTTTTTTGCAGCAGGCAATCAGCTTCTGATTATTGTGGACTCAGATGTCAGT  
GATGCTGGGAAATACACATGTGAGATGTCTAACACCCCTTGGCACTGAGAGAGGAAACGTG  
CGCCTCAGTGTGATCCCCACTCCAACCTGCGACTCCCCTCAGATGACAGCCCCATCGTTA  
GACGATGACGGATGGGCCACTGTGGGTGTCTGATCATAGCCGTGGTTTGCTGTGTGGTG  
GGCACGTCACTCGTGTGGGTGGTCATCATATACCACACAAGGCGGAGGAATGAAGATTGC  
AGCATTACCAACACAGATGAGACCAACTTGCCAGCAGATATTCCTAGTTATTTGTCTCT  
CAGGGAACGTTAGCTGACAGGCAGGATGGGTACGTGTCTTCAGAAAGTGGAAGCCACCAC  
CAGTTTGTCTCATCTTCAGGTGCTGGATTTTTCTTACCACAACATGACAGTAGTGGGACC  
TGCCATATTGACAATAGCAGTGAAGCTGATGTGGAAGCTGCCACAGATCTGTTCCCTTTGT  
CCGTTTTTGGGATCCACAGGCCCTATGTATTTGAAGGGAAATGTGTATGGCTCAGATCCT  
TTTGAAACATATCATACAGGTTGCAGTCTTGACCCAAGAACAGTTTAAATGGACCACTAT  
GAGCCAGTTACATAAAGAAAAAGGAGTGCTACCCATGTTCTCATCCTTCAGAAGAATCC  
TGCGAACGGAGCTTCAGTAATATATCGTGGCCTTCACATGTGAGGAAGCTACTTAACACT

FIG. 37B

101 / 108

AGTTACTCTCACAATGAAGGACCTGGAATGAAAAATCTGTGTCTAAACAAGTCCTCTTTA  
GATTTTAGTGCAAATCCAGAGCCAGCGTCGGTTGCCTCGAGTAATTCTTTCATGGGTACC  
TTTGGAAAAGCTCTCAGGAGACCTCACCTAGATGCCTATTCAAGCTTTGGACAGCCATCA  
GATTGTCAGCCAAGAGCCTTTTATTTGAAAGCTCATTCTTCCCCAGACTTGGACTCTGGG  
TCAGAGGAAGATGGGAAAGAAAGGACAGATTTTCAGGAAGAAAATCACATTTGTACCTTT  
AAACAGACTTTAGAAAACCTACAGGACTCCAAATTTTCAGTCTTATGACTTGGACACA[TAG]  
ACTGAATGAGACCAAAGGAAAAGCTTAACATACTACCTCAAGTGAACCTTTTATTTAAAAG  
AGAGAGAATCTTATGTTTTTTAAATGGAGTTATGAATTTTAAAAGGATAAAAATGCTTTA  
TTTATACAGATGAACCAAATTACAAAAAGTTATGAAAATTTTATACTGGGAATGATGC  
TCATATAAGAATACCTTTTTTAACTATTTTTTAACTTTGTTTTATGCAAAAAGTATCTT  
ACGTAAATTAATGATATAAATCATGATTATTTTATGTATTTTATAATGCCAGATTCTT  
TTTATGGAAAATGAGTTACTAAAGCATTTTAAATAATACCTGCCTTGTACCATTTTTTAA  
ATAGAAGTTACTTCATTATATTTTGCACATTATTTAATAAAATGTGTCAATTTGAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAA

FIG. 37C

102/108

SEQ ID NO:300

&gt;&lt;MW: 123434, pI: 6.09, NX(S/T): 12

MSAPSLRARAAGLGLLLCAVLGRAGRSDSGGRGELGQPSGVAAERPCPTTCRCLGDLDC  
SRKRLARLPEPLPSWVARLDLSHNRLSFIKASSMSHLQSLREVKLNNNELETIPNLGPVS  
ANITLLSLAGNRIVEILPEHLKEFQSLETLDLSSNNISELQTAFPALQLKYLYLNSNRVT  
SMEPGYFDNLANTLLVLKLNRRNRI SAIPPKMFKLPQLQHLELNRNKIKNVDGLTFQGLGA  
LKSLKMQRNGVTKLMDGAFWGLSNMEILQLDHNNLTEITKGWLYGLLMLQELHLSQNAIN  
RISPDWEFCQKLSELDLTFNHL SRLDDSSFLGLSLLNTLHIGNNRVSYIADCAFRGLSS  
LKTLDLKNNEISWTIEDMNGAFSGLDKLRLILQGNRIRSITKKAFTGLDALEHLDLSDN  
AIMSLQGNAFSOMKKLQQLHLNTSSLLCDCQLKWLPQWVAENNFQS FVNASCAHPQLLKG  
RSIFAVSPDGFVCDDFPKPQITVQPETQSAIKGSNLSFICSAASSSDSPMTFAWKDNE  
LHDAEMENY AHLRAQGGEVMEYTTILRLREVEFASEGKYQCVISNHFGSSYSVKAKLTVN  
MLPSFTKTPMDLTIRAGAMARLECAAVGHPAPQIAWQKGGTDFPAARERRMHVMPEDDV  
FFIVDVKIEDIGVYSCTAQNSAGSISANATLTVLETSPFLRPLLDRTVTKGETAVLQCIA  
GGSPPPKLNWTKDDSPLVVTERHFFAAGNQLLIIVDSVDSDAGKYTCEMSNTLGTERGNV  
RLSVIPTPTCDSPQMTAPSLDDDGWATVGVVIIAVVCCVVGTSLVWVVIIYHTRRRNEDC  
SITNTDETNPADIPSYLSSQGLADRQDGYVSSESGSHHQFVTSSGAGFFLPQHDSSGT  
CHIDNSSEADVEAATDLFLCPFLGSTGPMYLGKNVYGSDPFETYHTGCSDPDRTVLM  
DHYEPSYIKKKECYPCSHPSEESCERSFSNISWPSHVRKLLNTSYSHNEGPGMKNLCLNKSS  
LDFSANPEPASVASSNSFMGTFGKALRRPHLDAYSSFGQPSDCQPRAFYLKAHSSPDLD  
SGSEEDGKERTDFQEENHICTFKQTLNRYRTPNFQSYDLDT

**FIG. 38**

103 / 108

>1 A58532 glial cell membrane glycoprotein IIG-1 precursor - mouse (1091 aa)  
Score = 2619 (921.9 bits), Expect = 2.5e-274, Sum P(2) = 2.5e-274  
Identities = 500/836 (59%), Positives = 639/836 (76%), at 382,42, Frame = +1

DNA37140	382	PCPTTCRCGLDLDCCSRKRLARLPEPLPSWVARLDLSHNRLSFIKASSMSHLQSLREVKL ** * * *. **** . ** ** **** *.***.**** * ... * ..** *
A58532	42	PCAAACAGNSLDCSGRGLATLPDRDLPSWTTRSLNLSYNRLSEIDSAAFEDLTNLQEVL
DNA37140	562	NNNELETIPNLGPVSANITLLSLAGNRIVEILPEHLKEFQSLETLDLSSNNISELQTA-F *,*** **,** * . * * * *.** . ** . *** ******.***** *
A58532	102	NSNELTAIPSLGTASIGVVSFLQHKNILSVDSGSQLKSYLEVLDLSSNNITEIRSSCF
DNA37140	739	P-ALQLKYLYLNSNRVTSMESPGYFDNLANTLLVLKLNRRNRISAIPPKMFKLPQLQHLELN

**FIG. 39A**



104/108

```

*... * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
162  PNGLRIRRELNLASNRISILESGAFDGLSRSLTLRLSKNRITQLPVPKAFKLPRLTQLDLN
DNA37140
916  RNKIKNVVDGLTFQGLGALKSLKMQRNGVTKLMDGAFWGLSNMEILQLDHNNTLTKGWL
**.. ..***** **..***** **..***** **..***** **..***** **..*****
222  RNRIRLIEGLTFQGLDSLEVLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVNSGSL
DNA37140
1096 YGLLMLQELHLSQNAINRISPDWAEFCQKLSELDLTFNHLSRLDDSSFLGLSLNTLHIG
*** *..***** **..***** **..***** **..***** **..***** **..*****
282  YGLTALHQLHLSNNSISRIQRDGSFCQKLHELILSFNNLTRLDEESLAEISSLSILRLS
DNA37140
1276 NNRVSYIADCAFRGLSSLKTLDLKNNNEISWTIEDMNGAFSGLDKLRRLILQGNRIRITK
*..***. ***. ***. ***. ***. ***. ***. ***. ***. ***. ***. ***. ***.
342  HNAISHIAEGAFKGLSLRVLDDHNEISGTIEDTSGAFTGLDNLKLTFLGKNKIKSVAK
DNA37140
1456 KAFTGLDALEHLDSLDAIMSLQNAFSQMKKQLQLHLNTSSLLCDCQLKWLPQWVAENN
*..***.***** **..***.***.***.***.***.***.***.***.***.***.***.***.
402  RAFSGLESLEHLNLGENAIRSVQFDFAKMKNLKELYISSESFLCDCQLKWLPWLMGRM
DNA37140
1636 FQSFVNASCAPQLLKGRSIFAVSPDGFVCDFFPKPQITVQPETQSAIKGSNLSFICSAA
**..***.***.***.***.***.***.***.***.***.***.***.***.***.***.***.***
462  LQAFVTATCAHPESLKGQSIFSVLPDSFVCDFFPKPQIITQPETTMVAVGKDIRFTCSAA
DNA37140
1816 SSSDSPMTFAWKKNELLHDAEMENYAHLRAGGGEVMEYTTILRLREVEFASEGKYQCVI
*** ***** **..***.***.***.***.***.***.***.***.***.***.***.***.
522  SSSSPMTFAWKKNEDVLNADMENFAHVRAQDGEVMEYTTILHLRHVTFGHEGRYQCII
DNA37140
1996 SNHEGSSYSVKAKLTVNMLPSFTKTPMDLTIRAGAMARLECAAVGHPAFQIAWQKDGTD
***** **..***.***.***.***.***.***.***.***.***.***.***.***.***.
582  TNHFGSTYSHKARLTVNVLPSFTKIPHDIAIRTGTARLECAATGHPNPQIAWQKDGTD

```

FIG. 39B

105/ 108

```

DNA37140 2176 FPAARERRMHVMPEDDVFFIVDVKIEDIGVYCTAQNAGSISANATLTVLETPSFRLPL
A58532 642 FPAARERRMHVMPDDDDVFFITDVKIDDMGVYCTAQNAGSVSANATLTVLETPSLAVPL

DNA37140 2356 LDRTVTKGETAVLQCIAGGPPPKLNWTKDDSPLVVTERHFFAAGNQLLIIVDSVDSDAG
A58532 702 EDRVVTVGETVAFQCKATGSPPTPRITWLKGRPLSLTERHHFTPGNQLLVVQNVMMIDDAG

DNA37140 2536 KYTCMSNTLGTGERGNVRLSVIPTCDSPQMTAPSLDDDGWATGVVVIIVVCCVVGTS
A58532 762 RYTCEMSNPLGTERAHSQLSILPTPGCRK-----DG-TTVGIFTIAVVCISIVLTS

DNA37140 2716 LVWVVIIYHTRRRNEDCSITNTDETNPADIPSYLSSQGLADRDQGVVSSSGSHHQ
A58532 811 LVWVCIIYQTRKKSEEYSVTNTDETI VPPDPSYLSQGLSDRQETVVRTEGG--HQ

```

Score = 49 (17.2 bits), Expect = 2.5e-274, Sum P(2) = 2.5e-274  
 Identities = 16/64 (25%), Positives = 23/64 (35%), at 2764,931, Frame = +1

```

DNA37140 2764 CSITNTDETNPADIPSYLSSQGLADRDQGVVSSSGSHHQFV-----TSSGAGFFL
EQ ID NO:302 A58532 931 CSDCSTDTAYHPQPVPRDSGQPGTASSQELRQHDREYSPHPYSGTADGSHTLGGSLYP

DNA37140 2923 PQHD
A58532 991 SNHD

```

Score = 42 (14.8 bits), Expect = 1.4e-273, Sum P(2) = 1.4e-273  
 Identities = 13/33 (39%), Positives = 14/33 (42%), at 3388,920, Frame = +1

FIG. 39C

106 / 108

DNA37140 3388 PHLDAYSSFGQPSDCQPRAFYLKAHSSP-DLDSG  
\*\* \*.\* \*\*\* \* \* \* \*\*\*  
SEQ ID NO:303 A58532 920 PHTTAHSGSAVCSDCSTDY---HPQPVPRDSG

FIG. 39D

107 / 108

SEQ ID NO:304	OLI1375 (33780.f1) ACTCCAAGGAAATCGGATCCGTTT
SEQ ID NO:366	OLI1376 (33780.p1) GCCTTCACTGGTTGGATGCATTTGGAGCATCTAGACCTGAGTGACAACGC
SEQ ID NO:305	OLI1377 (33780.r1) TTAGCAGCTGAGGATGGGCACAAC

FIG. 40

108/ 108

&gt;&lt;DNA33780: 2779257 OVARTUT03 gl545807 gb98rodp 33 -16&gt;

GAT

&gt;&lt;33780.f1 {underline=1-24, dir=f}&gt;

SEQ ID NO:307

ACTCCAAGGAAATCGGATCCGTTCTATTACTAAAAAA

&gt;&lt;33780.p1 {underline=1-50, dir=f}&gt;

SEQ ID NO:309

GCCTTCACTGGTTTGGATGCATTGGAGCATCTAGACCTGAGTGACAACGCAATCATGTCT

TTACAAGGCAATGCATTTTCACAAATGAAGAACTGCAACAATTGCATTTAAATACATCA

AGCCTTTTGTGCGATTGCCAGCTAAAATGGCTCCCACAGTGGGTGGCGGAAAACAACCTT

CAGAGCTTTGTAAATGCCAGTTGTGCCCATCCTCAGCTGCTA

&gt;&lt;33780.r1 {underline=1-24, dir=b}&gt;

SEQ ID NO:308

AAAGGAA

**FIG. 41**